Searchable multi-dimensional Data Lakes supporting Cognitive Film Production & Distribution for the Promotion of the EuropeaN Cultural HeritagE

Grant Agreement No 101095303

DELIVERABLE 2.5:

State-of-the-art analysis and Specification of SCENE solutions

Work Package: 2

LEAD BENEFICIARY:

HYPERTECH

Delivery Date: 30.07.2024

Project acronym	**SCENE**
Project full title	Searchable multi-dimensional Data Lakes supporting Cognitive Film Production & Distribution for the Promotion of the EuropeaN Cultural HeritagE
Programme	Horizon Europe
Topic	HORIZON-CL2-2022-HERITAGE-01-06
Type of Action	HORIZON-Research & Innovation Actions
Grant Agreement	101095303
Start day	1 February 2023
Duration	36 months

LEGAL NOTICE
This project has received funding from the European Union Horizon Research and Innovation programme under grant agreement No 101095303. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use, which might be made, of the following information. The views expressed in this report are those of the authors and do not necessarily reflect those of the European Commission.

Funded by the European Union

Document Information

Deliverable number	D2.5
Deliverable name	State-of-the-art analysis and Specification of SCENE solutions
Lead beneficiary	HYPERTECH
WP	2
Related task(s)	2.4
Type	Document, Report
Reviewers (Organisation)	ADDMA, WR
Delivery date	30.07.2024
Main author(s)	HYPERTECH
Contributor(s)	CERTH, LINKS, DTT, MOG, UPV, AUTH, EPICA, GOF, CETMA, FOKUS

Dissemination level

PU	Public	X
SEN	Sensitive, limited under the conditions of the Grant Agreement	
Classified R-UE/EU-R	EU RESTRICTED under the Commission Decision No2015/444	
Classified C-UE/EU-C	EU CONFIDENTIAL under the Commission Decision No2015/444	
Classified S-UE/EU-S	EU SECRET under the Commission Decision No2015/444	

Funded by the European Union

Document history

Version	Date	Changes	Reviewer/Contributor
V0.1	05/06/2024	Table of Contents	HYPERTECH
V0.2	10/7/2024	Collected contributions from partners	CERTH, LINKS, DTT, MOG, UPV, AUTH, EPICA, GOF, CETMA, FOKUS
V0.3	19/7/2024	Send for review	HYPERTECH
V0.4	25/7/2024	Updated based on reviewers' comments	HYPERTECH
V1.0	26/7/2024	Final version	HYPERTECH

Table of Contents

Funded by the European Union

Funded by the European Union

List of Figures

List of Tables

List of definitions & abbreviations

Abbreviation	Description
AI	Artificial Intelligence
WP	Work Package
D	Deliverable
T	Task
VoD	video on demand
DNN	Deep Neural Networks
YEAR	Young Audience Award
EFI	European Film Industry
AWS	Amazon Web Services
MPEG-7	Moving Picture Experts Group
DOLCE	Descriptive Ontology for Linguistic and Cognitive Engineering
COMM	Core Ontology for Multimedia Annotation
YAGO	Yet Another Great Ontology
DAM	Digital Assessment Management
AEM	Adobe Experience Manager
VR	Virtual Reality
AR	Augmented Reality
ToF	Time-of-Flight
TDoA	Time Difference of Arrival
PDoA	Phase Difference of Arrival
UWB	Ultra-WideBand
RTLS	Real-Time Location System
TDMA	Time-Division Multiple Access
IMU	Inertial Measurement Unit
MILC	Media Industry Licensing Content
LRML	Latent Relational Metric Learning via Memory-based Attention for Collaborative Rank-ing
DPCML	The Minority Matters: A Diversity-Promoting Collaborative Metric Learning Algorithm
SW	Software requirements
HW	Hardware requirements

Funded by
the European Union

Executive Summary

This deliverable is related to the state-of-the-art analysis conducted for the components of SCENE platform. According to the requirements collected from D2.1 and D2.2, the specifications of SCENE platform have been extracted, and the components that will be implemented to include each step of the film-making process have been defined. Upon this, and in order to create a more thorough and well-established architecture for the platform, an initial analysis of the software and hardware requirements of each of platform’s components has been conducted and is presented in this deliverable.

In addition, during T2.4 – “Technology exploration and H/W & S/W requirements”, an initial search has been conducted on the relevant to SCENE projects that are available. The aim of this search was to identify the scope of each related project, its outcomes, and if and how it could be leveraged from the SCENE project. Initial dissemination activities have been performed with the CresCine and REBOOT projects, while discussions have been initiated with Premiere project for common exploitation activities between the two projects.

1 Introduction

The SCENE project is primarily designed to provide the means for a modern and globally competitive European filmmaking industry by relying on two solid pillars, namely semantically cognitive AI technologies and (in)tangible European cultural assets, while always adhering to European values and policies regarding the human and the environment. SCENE will create a diverse range of cutting-edge technologies and services capable of covering the whole pipeline of the film-making business, "from farm to fork," that is, from the very early phases of film planning to their playback on the audience screen. Specifically, SCENE aims to innovate by enriching the existing data lakes with high-qualitative & editable 3D digital models of European cultural sites, and by enhancing their accessibility with the integration of multi-dimensional knowledge graphs carrying location- & cultural-aware information.

SCENE will further facilitate the film-making & film-editing side during the production and the post-production phase with efficient simulation & novel assessment mechanisms. Similarly, it will enable smart & privacy-preserving interaction channels between industrial stakeholders & the audience, that will allow not only for pro-active sensing of the audience's preferences in the pre-production phase, but also for the early identification of the most matching advertisement & distribution channels during the distribution phase, as well as the insightful matching & recommendation between films & individuals. The integration of the aforementioned technologies under the unique SCENE platform demonstrates significant capacity to render an excellent proof-of-concept prototype to guide, inspire and mark significant breakthrough in the filmmaking industry during the post COVID era, on scientific, commercial, business & policy level.

Therefore, keeping up with the latest techniques and technologies is essential for keeping a competitive advantage in the quickly changing filmmaking industry. To achieve this and gain valuable insights into the present technological landscape, a literature review on the technologies available in each field (e.g., Artificial Intelligence (AI), Blockchain, Computer vision, etc.) is essential, along with an evaluation of the cutting-edge technologies and current outcomes of similar projects to SCENE. By identifying best practices, comparing performance to rivals, and assisting in the decision-making process regarding technological investments, this assessment guarantees that the tools implemented within the project utilise the cutting-edge and efficient technologies available in literature.

Furthermore, the identification of the state-of-the-art technologies will ensure the seamless development and integration, customization, and scalability of various technologies required by each of SCENE's tools. Detailed state-of-the-art analysis facilitate interoperability, enhance efficiency, and future-proof technological investments, allowing for smooth collaboration and adaptation to new advancements.

1.1 Purpose, Context and Scope of the document

The purpose of this deliverable (D2.5 – State-of-the-art analysis and Specification of SCENE solutions) is to present the assessment of the state-of-the-art performed, including existing results and technologies relevant to SCENE, and specification of system modules.

Particularly, the first core part of this document outlines the assessment performed on the relevant to SCENE's projects identified, along with the assessment of the state-of-the-art and existing results from related projects to identify which are relevant to SCENE. This deliverable presents also an analysis of the priority and applicability of the explored technologies in relation to the SCENE platform, along with a state-of-the art analysis with respect to the scientific fields required for the implementation of the platform. Finally, this deliverable aims to capture and present the software and hardware requirements of all SCENE platform's components, along with any existing limitations, resulting from the requirements collected during T2.2 – End-user needs and Requirements.

1.2 Relationship with other tasks and deliverables

The present deliverable is related with tasks from WP2, WP3 and WP4, which are relevant to the requirements collection and SCENE's tools implementation. More specifically, T2.4 – "Technology Exploration and hardware and Software requirements" is related to T2.1 – "Use Cases Definition & Application Scenarios" because this task and its corresponding deliverable 2.1, presenting the use cases defined for each of the tools described in SCENE project, and setting the basis for the initial analysis of the technologies required for each use case and tool. Furthermore, T2.4 is related to T2.2 – "End-User needs & Requirements" since this task is responsible for the requirements collection and analysis of each tool. Based on the collected requirements, the software and hardware specifications are extracted within T2.4, and are described in this deliverable. The results of this task (T2.4) including the literature review analysis conducted along with the software and hardware specifications defined will be used as input in T2.5 – "System Specification & Architecture Definition".

For WP3 and WP4, T2.4 is related with T3.1 – "Exploration Data Lakes & Ontologies", T4.1 – "Location Scouting", T4.2 – "AI-based Audience Preferences Scouting Tool", T4.3 – "Audience Building", T4.4 – "Lighting & Audio Simulation tools", T4.5 – "Post Production Effects", and T4.6 – "Distribution Engine & Recommendation System", respectively. Task 2.4 captures the scope and implementation requirements and efforts of each of the tools developed within the aforementioned WPs, and the literature review conducted for each of these tools, in order to extract software and hardware specifications for each tool. These extracted specifications will be lever-aged by T5.1 – "Platform Integration" and T5.3 – "Pilots Installation and Execution" in order to set up the pilot environments properly according to the needs of the SCENE platform, while facilitate the integration of each tool into the common SCENE platform (within T5.1). At last, T2.4 is related to T6.2 – "Market Analysis, Business Model Definition and Exploitation Strategy Tool landscape analysis", more specifically under the tool landscape analysis conducted, that was aligned with captures also the technologies developed and used by each of SCENE's tools and aimed at capturing an image of was technologies are being commonly used by the competition.

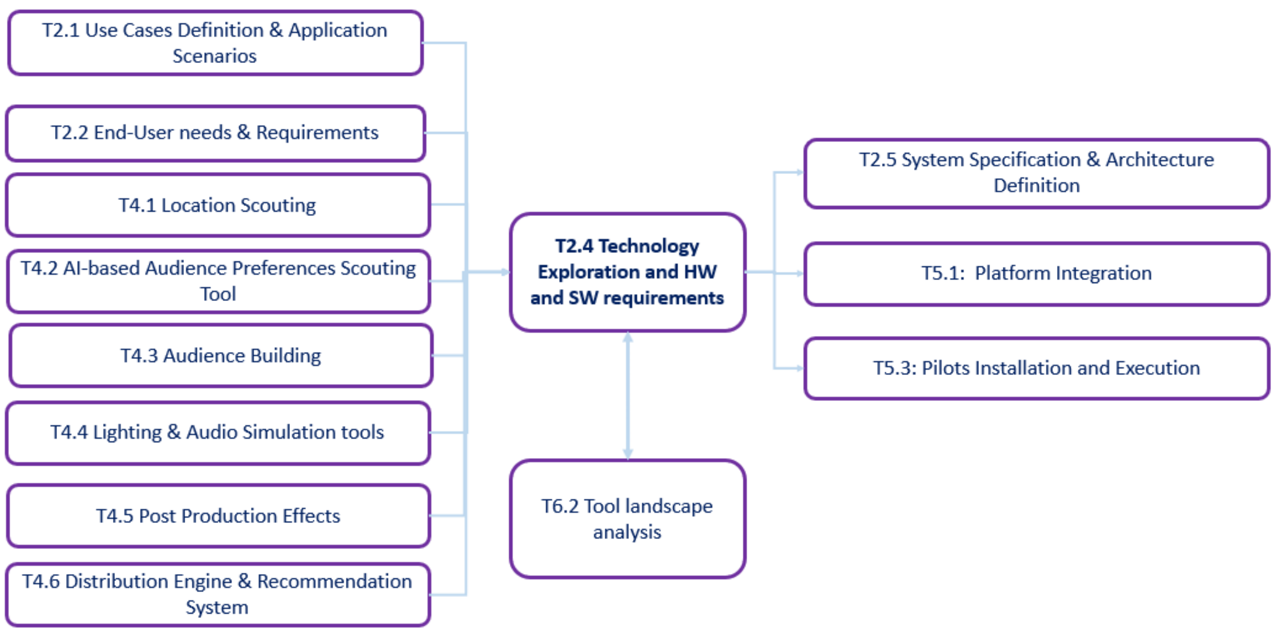


Figure 1: Relations of T2.4 and its corresponding D2.5 with other tasks

Funded by the European Union

1.3 Structure of the deliverable

This deliverable is structured and organized as follows. Section 1 provides the scope of this deliverable in detail, its relation with other deliverables and a brief description of the deliverable's structure. Section 2 presents the results of the projects that are relevant to SCENE project, highlighting their functionalities, innovation and information that could be leveraged by SCENE. Section 3 presents the state-of-art analysis conducted for each one of the tools that will be implemented within SCENE, along with the technologies developed including Artificial Intelligence (AI), blockchain, image processing, 3D reconstruction, etc. Section 4 includes the software and hardware specifications extracted per tool, along with any requirements and limitations resulted after the collection of the requirements collected within D2.2 and D2.3 and the first version of the SCENE architecture's implemented within D2.6. Finally, section 5 is a conclusion presenting the outcomes of the analysis conducted within T2.5 and the next steps of the project.

2 Projects relevant to SCENE

The scope of this section is to identify the projects that are relevant to either the SCENE project or the technologies implemented in the SCENE project, and analyse their relevancies with the SCENE project. Some of the identified as similar projects are currently running, while others have already finished. The analysis of each project includes the assessment of its results, final or current, in terms of:

- Functionality provided
- Innovation capacity
- Technology
- License

For each one of the projects, a description of its objective and results is provided, along with a detailed presentation on the aforementioned terms.

2.1 The MovieLabs Creative Works Ontology Project

MovieLabs Creative Works Ontology Project[1] is an initiative by MovieLabs, an organization founded by major Hollywood studios to advance technologies for the motion picture and television industries. This project aims to create a standardized and comprehensive ontology for representing various aspects of creative works, particularly in film and television.

The ontology is designed to facilitate interoperability and data exchange between different systems and organizations within the media and entertainment industry. It covers a wide range of concepts and relationships relevant to the lifecycle of creative works, including:

1. **Content Creation**: Information about the production process, such as scripts, storylines, characters, and crew.
2. **Content Distribution**: Data related to the distribution and licensing of creative works, including rights management and delivery formats.
3. **Content Consumption**: Details about how and where content is consumed, such as viewing platforms, audience engagement, and analytics.

The goal of the Creative Works Ontology is to provide a common framework that can be used to describe and manage media content more efficiently, enabling better collaboration, automation, and innovation across the industry. This ontology helps streamline processes, reduce redundancy, and improve the accuracy and consistency of data handling in the creation, distribution, and consumption of media content.

The SCENE project will leverage the MovieLabs ontology, and use the included concepts and axioms as a baseline, aiming to extend it based on the requirements collected within the SCENE project and the tools that will be implemented, capturing the current needs of the filmmaking industry. The SCENE project will leverage the available ontologies from literature, and will use the entities relevant to the project and the film making industry's needs, enhanced with location, regional and cultural-aware information in order to create an extended ontology. Additional information on this extension is provided in sub-section 3.1.2.

2.2 The Semantic Web for Audio-Visual Content Project

The goal of the project was to create a research group on sophisticated audio processing for multimedia management at the University of Žilina, Slovakia[2] [69]. By creating audio analysis tools that can extract knowledge and semantics from digital audio-visual content, this programme aimed to advance the state-of-

[1] https://mc.movielabs.com/docs/ontology/ (Accessed on 24 July 2024).

[2] Teerapong Serisamran, "The analysis of film editing techniques in Thai best editing feature film "Bad Genius" International Journal of Management and Applied Science, ISSN: 2394-7926 Volume-5, Issue-3, Mar.-2019, http://www.iraj.in/journal/journal_file/journal_pdf/14-551-15599674374-7.pdf (Accessed on 24 July 2024).

Funded by the European Union

the-art in the field of multimedia processing and semantic analysis, which is crucial for enabling quick, efficient, and user-friendly access to digital multimedia resources including Internet audio and video, digital libraries, and television archives. Additionally, it makes content delivery easier for a variety of media platforms, such as mobile and handheld devices. Combining aural and visual clues can clear up ambiguities and produce more accurate results when either one of the two types of information is insufficient to determine the content of the scene.

In addition, a new technique has been developed within this project for signal parameterization in voice analysis and recognition. This technique produces a compact representation of speech in the time-frequency domain by using two-dimensional cepstral analysis. When compared to conventional techniques, it successfully minimises the quantity of voice features, solving the critical issues of memory and computing efficiency in multimedia content management. Key-word detecting algorithms can incorporate this new method. In many applications, such as audio and video material retrieval, subject detection, and genre identification, keyword spotting is crucial.

To facilitate combined audio and visual analysis, an MPEG-7-based browsing tool was created. By the project's end, this tool could analyse metadata representing the temporal breakdown of audio and video content. It was created primarily for intelligent video navigation, with capabilities including playing video shots or audio samples from current material, organising important frames of video segments in a scroll grid, and painting a timeline. This tool proved the usefulness of the MPEG-7 content description interface standard. It had a modular architecture that allowed for future extension and could be used for demonstrations, testing, and assessment.

The results of the analysis on multimedia conducted within this project will be leveraged by the audio simulation tool implemented within WP4 in SCENE project. Also, the information required as input for the MPEG-7 tool and the data resulted from the multimedia analysis have been captured by the SCENE-O ontology implemented within WP3.

2.3 The DEEP FILM Access Project: Ontology and Metadata Design for Digital Film Production Assets

The DEEP FILM Access Project (DFAP)[3] aims to tap into the hidden potential within the vast and complex data sets generated by industrial digital film production. The complexity of filmmaking has now reached a level where a new on-set role, 'data wrangling,' has emerged to manage camera-generated data. This task is further complicated by the integration of Computer-Generated Imagery and Stereoscopic 3D filming. Even small independent films often use multiple cameras, which significantly increases the volume of raw material and the complexity of data interrelations. Additionally, data produced by the creative process, such as director and crew notes on quality, logistics, shot organization, and props, is recorded separately from camera data. Current archival methods for this data are redundant and prone to errors, making it difficult to search across different data types in an integrated manner. As filmmaking transitions from photo-chemical to digital, new archival methods are necessary to handle this data, offering the potential for novel and in-depth analysis of the filmmaking process and its outcomes.

An integrated process and a relevant framework were created by the DFAP for the management of all of the assets created by digital feature film production. The first objective is the design classifications and definitions that will standardise the description, layering and interlinking of data assets making them accessible online. Next, a method will be developed to integrate this information with records from everyone involved

[3] https://gtr.ukri.org/projects?ref=AH%2FL010305%2F1 (Accessed on 24 July 2024).

Funded by the European Union

in the film's production (actors, crew, etc.). This integration will allow for joint interrogation of camera-generated and creatively generated data. The questions that this joint interrogation opens up will be explored by the DFAP project.

The DFAP project recognizes the critical need to develop comprehensive protocols and best practices for indexing and accessing film assets and metadata for research and teaching. The project's goal is to make it easier to interrogate the many data created during the film production process by inventing a novel technique to searching metadata. Although metadata cannot represent the entire breadth of the data, our cross-disciplinary and multi-partner methodology aims to protect and improve the core dataset's integrity through a systematic and thorough modelling approach.

As a result, the Deep Film Access Project has created a semantic infrastructure that aids in the integration of data and metadata generated throughout the feature-film production lifecycle. The Deep Film Access Project laid the groundwork for a knowledge architecture that enables the automated administration of feature film digital assets. This is based on a process analysis using an OWL ontology.

2.4 The Moving Cinema project

Moving Cinema is a European project launched in September 2014 with the help of MEDIA Creative Europe. More than 32,000 young people (aged 12 to 19), 50 organisations, 120 filmmakers, 1.350 instructors, and 220 schools have participated in this initiative.

This project aims to foster strong bonds between young people and film by providing to them the tools they need to be independent viewers, and, eventually, to produce active and sensitive audiences capable of appreciating a wide range of cinematographic expressions. This project specialises on modern and historic European auteur cinema, with a particular emphasis on films and directors that deviate from the mainstream while still engaging young people in a profound and meaningful way.

In its five years, Moving Cinema explored five strands of work: (1) attendance at festivals and screenings; (2) young film programmers; (3) filming with mobile devices inspired by great European filmmakers; (4) access to films on video on demand (VOD) platforms on their own, so that they may retain independent relationships with European films; and (5) Inside Cinema, an online platform that documents and provides valuable materials for anyone interested in the creative process of films. In addition, they have also delivered training programmes in the partner organizations' home countries as well as other European nations.

It also develops approaches and tactics that may be applied in other circumstances. Each activity is then analyzed in order to establish functioning models and their underlying processes. Educational materials have been created that are now available on the Internet and are meant to benefit all organizations, institutions, and individuals interested in the transmission of film, including the creation of a catalogue of European films chosen and programmed since 2016 by the Young Programmers.

This catalogue[4] has been leveraged by the SCENE project, where its movies have been enclosed into the knowledge base of SCENE. For each film, information and different types of materials have been collected, including the following:

- Presentation texts, videos, trailers, posters, flyers, booklets, plastic creations and selections of quotes by filmmakers or film crew.
- Written and visual synopsis.
- Indications regarding how to access the film: contact details of distributors and/or VoD platforms, available subtitles, etc.
- Links with the 'FilmViewing-FilmMaking Kits' connected with each film.
- Possible subjects and connections with the school curriculum.

[4] https://movingcinema.eu/resources/young-programmers-catalogue/

Funded by
the European Union

2.5 The Premiere project

The PREMIERE concept[5] represents the start of a new era in the performing arts. Using cutting-edge AI and XR technologies, it will permit the whole performance life cycle, from production to diffusion and preservation. Research institutions and creative partners are pushing the development of digital tools to improve archives, expand venues as virtual platforms for live performances and co-creation, and innovate at the intersection of art and technology. Overall, PREMIERE stands for accessible access, competitive CCIs, critical digital involvement, and cultural heritage conservation.

Premier is a project that has been running since 2022 and is projected to be completed in 2025. It has cooperated with the University Jean Monnet to create complete video processing and 3D reconstruction capabilities. These technologies include 2D and 3D posture estimates, tracking, and trajectory analysis. The research uses advanced computer vision algorithms and AI, namely Deep Neural Networks (DNN), to assess and model modern dance films. They are also developing 3D models for static objects in audio-visual archives with the goal of generalising these technologies to other types of audio-visual content, such as theatre.

The goals include:

- Scene Analysis and Understanding by creating a dataset of live performances representatives of complex dance or theatre events using several cameras from different points of views.
- 3D pose trajectories estimation in complex scenes by building a 3D model of moving people in these audio-visual contents from multiple views. Including multi-people pose and motion estimation, segmentations and tracking methods and action parsing methods.
- Creating 3D models for static elements in audio-visuals archives with object detection, segmentation, and tracking.
- Create interactive platforms that leverage digital technologies AR/VR to address theatre's challenges of intangibility and ephemerality, broadening audience reach, fostering new artistic expressions, and enhancing accessibility.
- Premier with cutting edge technologies, offers:
- Performing arts archives browsing enhanced with AR/VR technologies.
- Live performance enhanced with VR technologies.
- Actor/dancer virtual co-creation performance.
- Dance-based artistic creation environment.

The SCENE project will make a synergy with the Premiere project on two main pillars: (a) discuss about the introduction of the 3D models created within the Premiere project into the Audience Building tool implemented within SCENE project, and (b) to leverage the scene and performance analysis results conducted within the Premiere project, to enhance the ontology and the metadata collected for each film within SCENE project. Also, SCENE is able to provide to Premiere information about the 3D models created from cultural environments in Italy and Cyprus, while provide their knowledge gained from the use and evaluation of SCENE platform.

2.6 The REBOOT project

The REBOOT project[6] aims to revive, boost, optimise, and transform European film competitiveness. This project focuses on leveraging existing strengths, fixing weaknesses, and proactively planning for the European film industry's future competitiveness across all regulatory, practical, and experiential aspects. The project's goal is to investigate audience preferences and generational dynamics in order to gain a thorough knowledge and increase young people's involvement with European cinema. Furthermore, the initiative aims

[5] https://premiere-project.eu/the-project/ (Accessed on 26 July 2024).
[6] https://thereboot-project.eu (Accessed on 26 July 2024).

to strengthen the European Union's position in the global audio-visual market and encourage cultural diversity in the European film industry.

The five dimensions of the project includes:

- Increasing support for young people's engagement with European Film.
- Strengthening the place of the EU in the global audio-visuals' economy.
- Supporting cultural diversity in the EU film industry.
- Addressing the need for a different understanding of competitiveness.
- Recognizing and supporting the importance for the EU of film and, more broadly, of the cultural and creative sector as a geopolitical asset.

The European film industry is hampered by educational deficiencies and structural barriers, such as a lack of structured film education programmes, resulting in limited film literacy among young people, as well as insufficient educational opportunities and entry into the industry for prospective professionals. European youth prefer US films, perceiving European works as more cerebral, artistic, and less entertaining, a choice aided by an infrastructure that favours Hollywood productions while restricting the distribution of national and European films. Although efforts such as the Young Audience Award (YEAR) and pre-industry programmes seek to involve young people and develop cinema literacy, their effectiveness is limited by low exposure and accessibility. Involving young people in decision-making procedures within the film business is critical to keeping the relevance and dynamism of European cinema.

The COVID-19 epidemic has had a tremendous impact on the worldwide audio-visuals market, with US-based streaming services like Netflix and Disney Plus strengthening their economic supremacy and increasing their soft power. This position creates both obstacles and possibilities for the EU, which has responded with regulatory measures such as the AVMSD's quota for European content to foster cultural diversity. A more comprehensive view of competitiveness that takes into account cultural and social repercussions is required, as is recognition of the cultural sector's strategic value as a geopolitical asset. Furthermore, there is potential for regional media platform efforts in reaction to the dominance of US platforms, emphasising the importance of the EU supporting and adapting its audio-visuals industry's competitiveness.

The worldwide popularity of non-Western TV dramas, which combine local and global components and draw on distinct cultural narratives, shows how such techniques may greatly increase international reach. This implies that the EU may pursue similar strategies to boost its audio-visual sector and compete more successfully on a global basis. By encouraging cultural variety and displaying a diverse spectrum of cultural expressions, these dramas emphasise the relevance of cultural resonance and audience involvement in competitiveness, which supports both economic growth and cultural impact. Recognising film's position as a strong instrument of cultural diplomacy, these triumphs strengthen their respective countries' soft power.

By the end of the project, REBOOT aspires to create a multilevel, comprehensive, and evidence-based set of data to assist with planning for European film success across the supply chain. The results will include a comprehensive set of policy and action recommendations for all stakeholders, derived from experimental and empirical comparative research across the continent and beyond. Additionally, REBOOT will compile datasets with empirical data from thousands of interviews and survey responses from audiences, as well as datasets on the law and governance of the European film sector. Furthermore, REBOOT will outline clear pathways for the development of an inclusive, innovative, culturally diverse, democracy-committed, and sustainable European Film Industry (EFI).

Due to the participatory design approach followed for the development of SCENE's tools, SCENE will leverage the guidelines and policy recommendations that REBOOT project will prepare. In addition, synergy between these two projects will be conducted in order to create a common dataset with the requirements collected on both aspects from the interviews and surveys conducted between the corresponding stakeholders of each project.

2.7 The CresCine project

CresCine project's[7] overall objective is to enhance the competitiveness and cultural diversity of the European film industry. This will be achieved by understanding, engaging with, empowering, and ultimately transforming European small markets through original research and piloting the results in Estonia, Lithuania, Denmark, Ireland, Belgium (Flanders), Croatia, and Portugal. The CRESCINE report collects data on the production, distribution, exhibition, and reception of films from seven selected small ecosystems in Europe, as well as from six large markets: France, Germany, Italy, Spain, the UK, and Poland.

Analyzing the production volume of small European film markets from 2014 to 2022, it has been shown that Denmark and Estonia maintain stability during COVID-19, while Croatia experienced declines. Ireland excels in co-productions, highlighting international ties. Dominant genres are drama and comedy, with Portugal and Ireland favoring drama, and Lithuania, Estonia, and Flanders producing more comedies. Denmark and Flanders also create significant family films.

CresCine investigates the film industries of the seven CresCine Ecosystems - Croatia, Denmark, Estonia, Flanders, Lithuania, and Portugal - utilising over 30 quantitative indicators (data sources) that cover the whole value chain, from production and funding to distribution, exhibition, and reception. It focuses on the unique competitive advantages and challenges that the film industry encounters in small markets. This research also summarises the obstacles that tiny film markets face in the production, funding, distribution, exhibition, and reception of feature films. Furthermore, the study compares small markets to larger markets, specifically in terms of big markets' view on competitiveness and export, film exports, and the influence of US films on local businesses on the areas of competitiveness between small and large markets. The findings show that none of the ecosystems have a strong enough domestic market to support a film industry on its own terms. At the same time, most markets (save Ireland and, to a lesser extent, Flanders) rely heavily on their home audience to obtain admissions. All ecosystems rely heavily on public financing and/or investment programmes, and thus struggle to secure or sustain private investment, such as from distributors, manufacturers, and other private investors. Distributors active in minor markets do not, in most circumstances, keep domestic titles in their portfolio.

Exports are not a strategic priority for the majority of the minor markets assessed. Volatility in finance options, production volume, admissions, and market shares are a prominent element in most of the analysed markets, which has been worsened by the Covid-19 epidemic, from which neither of the studied markets has recovered in terms of admissions as of 2023. The lesser output of films in smaller markets also limits the variety and frequency with which specific styles and genres may be supplied. Popular genres like as action and sci-fi are rarely provided domestically, instead being derived mostly from the United States, resulting in high admissions for US films. In most areas, the lack of many Hollywood blockbusters delayed by the epidemic increased domestic shares, indicating that the long-standing issue of dominance of US films continues to be significant. The SCENE project will leverage the outcomes of the CresCine project, studying the results of the film making industry collected by the CresCine partners. In addition, SCENE will share the knowledge extracted from the requirements collection and SCENE platform's evaluation phase to CresCine project, aiming to initiate a communication about its exploitation by the film-making industry.

At this point it should be mentioned that SCENE project has co-organised a workshop with CresCine and Reboot projects for the ethical approach of generative AI. Additional information on this workshop is available in D6.5. Also, an initial discussion with PREMIERE project has started on the next steps on their collaboration. For the rest of the projects and initiatives, SCENE will attempt to initiate relations either through our own growing networks or from participation in common events. The exploration of other projects related to SCENE will continue with the aim to further stimulate liaison and co-operation activities with other projects.

[7] https://www.crescine.eu/ (Accessed on 26 July 2024).

Funded by the European Union

3 State-of-the-art technologies analysis

In this section, the cutting-edge technologies that form the foundation of SCENE platform are presented. The rapid speed of technological change needs a careful evaluation of the most recent advancements to guarantee that SCENE platform remains competitive in matters of efficiency, security, and user satisfaction. By investigating these cutting-edge technologies, we want to gain a thorough grasp of their capabilities, integration procedures, and the value they provide to our system.

This analysis focuses on the present best practices, emerging trends, and prospective future advancements in the area of each tool. The aim of this analysis is not only to demonstrate the strength of SCENE platform, but also identify any need on utilising the most current technology available. In addition, this analysis will provide stakeholders with insights into the reasoning for the technology choices and how they work together to improve the SCENE platform's components performance and dependability.

3.1 Data Lakes and Ontologies

Data lakes have become essential in modern data management, offering a flexible and scalable solution for storing vast amounts of data. Unlike traditional data warehouses, data lakes allow raw data storage in its native format, which includes structured, semi-structured, and unstructured data. This summary provides an overview of the state of the art in data lakes, including their architecture, advantages, challenges, and recent advancements. A more detailed description of data lakes will be provided as part of task T3.1 in deliverable D3.1 (M22).

3.1.1 Data Lake Architecture

A typical (simplified) data lake architecture (Figure 2) consists of several key components:

1. **Ingestion Layer**: This layer is responsible for capturing and importing data from various sources, such as databases, real-time streams, and external APIs. Tools like Apache Kafka and AWS Glue are commonly used for data ingestion.
2. **Storage Layer**: Data is stored in its raw form in a scalable and cost-effective storage system, often using distributed file systems like Hadoop Distributed File System (HDFS) or cloud storage services like Amazon S3 and Google Cloud Storage.
3. **Processing/Analysis Layer**: This layer involves data processing and transformation using frameworks like Apache Spark, Apache Flink, or Presto. It allows for batch and real-time processing to prepare data for analysis.
4. **Governance Layer**: Ensures data quality, security, and compliance. It includes metadata management, data cataloguing, and access control. Tools like Apache Atlas and AWS Lake Formation provide governance capabilities.
5. **Application and Visualization Layer**: Provides tools for querying, analysing, and visualising data. SQL engines, machine learning frameworks, and BI tools like Tableau and Power BI are used in this layer. Sometimes, the border between the analysis and application layer is diffuse.

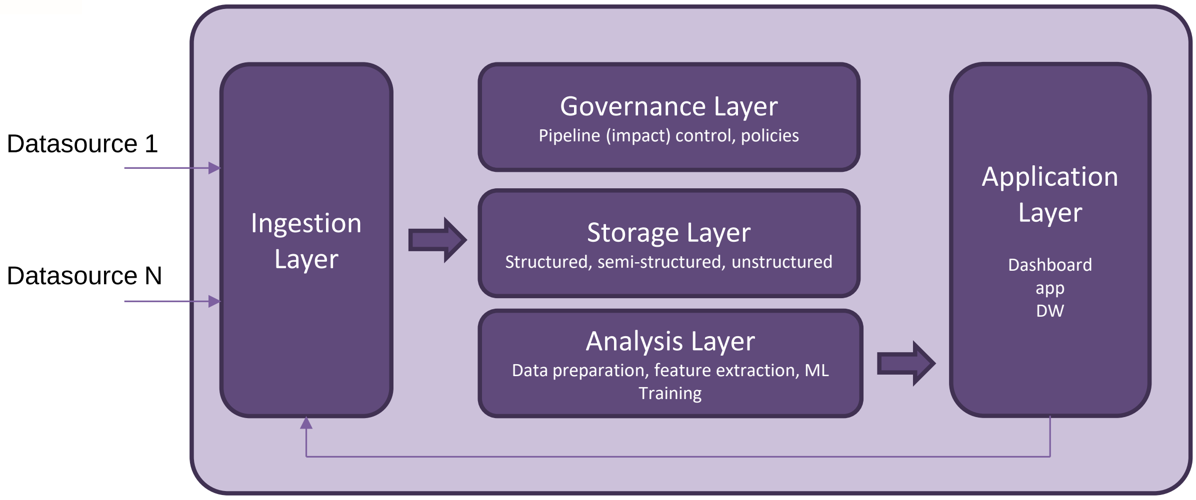


Figure 2. Data Lake simplified architecture

3.1.2 Challenges and advances in Data Lakes

The main advantages of using Data Lakes are the following:

1. **Scalability**: Data lakes can handle petabytes of data, making them suitable for large-scale data storage needs where the amount of information (e.g., continuous real-time data) occurs.
2. **Flexibility**: They support various data formats, including structured, semi-structured, and unstructured data. Data can be first ingested and, later on, processed/analysed.
3. **Cost-Effective**: Using commodity hardware or cloud storage reduces costs compared to traditional data (data warehouses)
4. **Advanced Analytics**: Data lakes enable advanced analytics, machine learning, and big data processing, as they intend to store all relevant data to promote the best insights.

From another perspective, data lakes encounter several challenges:

1- **Data Quality**: Ensuring data quality and consistency can be challenging due to the heterogeneous nature of the data. If not properly managed, data lakes can become *data swamps*.
2- **Data Governance**: Implementing robust frameworks to manage metadata, security, and compliance is critical. The Governance Layer is probably the most challenging aspect.
3- **Complexity**: Designing and maintaining a data lake architecture requires significant expertise and resources. Skilled workers must handle data management.
4- **Performance**: Query performance can be slower than data warehouses, especially for complex queries. This is a natural drawback when handling many different types of (unstructured) data.

Currently, research and implementation regarding data lakes are focussed on:

1. **Data Lakehouse**: A new architectural paradigm that combines the best features of data lakes and data warehouses. It supports ACID (atomicity, consistency, isolation, durability) transactions and provides a unified data management and analytics approach. *Delta Lake* is an open-source approach offering ACID transactions. Databricks, apart from the three major cloud providers, is probably the most relevant company focused on data Lakehouses.
2. **Serverless Data Lakes**: Cloud providers offer serverless data lake solutions that simplify the deployment and management of data lakes. Examples include AWS Lake Formation and Google Cloud Dataproc.
3. **Real-Time Data Lakes**: Integrating real-time data processing capabilities using technologies like Apache Kafka and Apache Flink to support streaming data ingestion and analysis.

4. **Enhanced Data Governance**: Advanced tools and frameworks for data governance, such as Apache Atlas and Microsoft Purview provide better metadata management, data lineage, and compliance tracking.

Regarding case studies focussed on the film-making industry, **Netflix** is supposed to use a data lake architecture to store and analyse vast amounts of streaming data for personalised content recommendations and operational analytics. However, the details of the approach are not publicly available.

3.1.3 Cloud-based and open-source approaches

The three major cloud providers—AWS, Google Cloud, and Microsoft Azure—offer robust data lake solutions tailored to diverse organisational needs. **AWS** provides Amazon S3 for scalable object storage, AWS Glue for data integration, and Amazon Athena for serverless querying, enabling comprehensive data lake management. **Google Cloud** features Google Cloud Storage for object storage, Google Dataflow for real-time data processing, and BigQuery for powerful analytics, facilitating seamless data lake operations. **Microsoft Azure** offers Azure Data Lake Storage for scalable storage, Azure Data Factory for data orchestration, and Azure Synapse Analytics for integrated analytics, providing a unified platform for efficiently building and managing data lakes. More information with sample diagrams will be provided in D3.1

From an open-source perspective, **MinIO[8], Dremio[9], and Ceph[10]** are powerful technologies that complement data lake architectures by enhancing storage, access, and data processing capabilities. **MinIO** provides high-performance, scalable object storage compatible with Amazon S3, making it an excellent choice for data lakes requiring fast and reliable storage solutions. It allows data lakes to leverage MinIO's ability to handle large-scale unstructured data, ensuring efficient data retrieval and storage operations. This option has been selected in SCENE as a core part of data storage for simplicity.

Dremio, on the other hand, offers a self-service data platform that enables interactive and fast queries across various data sources within a data lake. **Ceph**, commonly used in Kubernetes deployments, provides a unified, distributed storage system that supports object, block, and file storage. It also ensures high availability and fault tolerance, making it ideal for production environments, but it is more complex and requires advanced skills.

3.1.4 Ontologies

An ontology is a systematic framework for categorising and defining the relationships between ideas within a certain domain, resulting in a shared and common knowledge of that area. It serves as a formal representation of knowledge by defining entities, properties, and linkages, allowing for uniform data interpretation and interoperability across systems and applications. An ontology's value stems from its capacity to allow straightforward communication, increase data integration, and improve information retrieval by providing a cohesive structure that is consistent with the domain's unique vocabulary and conceptualizations. This is especially vital in complicated fields where exact and clear definitions are required for successful collaboration, decision-making, and process automation.

3.1.4.1 Challenges and advances in Ontologies

Despite their critical role in organising domain knowledge, ontologies suffer a number of major obstacles throughout development and implementation. One of the fundamental problems is producing a thorough

[8] https://min.io/ (Accessed on 26 July 2024).
[9] https://www.dremio.com/ (Accessed on 26 July 2024).
[10] https://ceph.io/en/ (Accessed on 26 July 2024).

Funded by the European Union

and accurate representation of complex domains, which frequently necessitates substantial domain expertise and coordination among multiple stakeholders. Furthermore, maintaining and updating ontologies to reflect developing knowledge within a subject can be time-consuming and need ongoing work.

Another significant difficulty is guaranteeing compatibility across several ontologies. Because many ontologies can be produced independently, integrating them into a cohesive framework might be difficult owing to terminology and structure differences. This lack of standardization might impede data exchange and collaboration across systems and organizations.

Despite these limitations, there have been significant breakthroughs in ontology creation and implementation. Automated tools and procedures for ontology development and refining have evolved, lowering the amount of manual labour necessary while enhancing accuracy. Machine learning and natural language processing techniques are being used to extract and organize information from enormous datasets, allowing for the production of more complete and dynamic ontologies.

Standardization initiatives, such as the creation of ontology languages like OWL (Web Ontology Language) and frameworks like RDF (Resource Description Framework), have considerably increased interoperability. These standards allow for the smooth integration and flow of information across many systems and domains. Furthermore, community-driven platforms and collaborative initiatives have facilitated the exchange of best practices and resources, furthering the discipline.

3.1.4.2 Available film-related ontologies in literature

Several ontologies have emerged in the field of cinema studies and the film business, with the goal of organizing and facilitating the huge and diverse knowledge in this area. These ontologies formalize film-related notions such as metadata, genres, actors, production information, and spectator interactions. Here, we examine some of the most common film-related ontologies and multimedia related ontologies in the literature, emphasizing their scope and contributions.

The available **multimedia ontologies** in literature are the following:

- Moving Picture Experts Group (MPEG-7) [1][9] [1][10]
  The MPEG-7 Ontology is a comprehensive framework designed to represent the rich set of multimedia content descriptions standardized by the Moving Picture Experts Group (MPEG). MPEG-7 provides a detailed and structured way to describe multimedia content, such as images, audio, video, and their various attributes. The ontology encompasses a wide range of entities, including descriptors for low-level features like color, texture, and shape, as well as high-level semantic entities like objects, events, visual features, audio features or more abstract concepts. By formalizing these descriptors within an ontology, MPEG-7 facilitates the creation of interoperable metadata, enabling advanced multimedia content management and retrieval.
  The scope of the MPEG-7 Ontology is broad, addressing the needs of various applications, from multimedia content indexing and retrieval to digital libraries and content-based multimedia search engines. Its primary objectives are to enhance the accessibility and usability of multimedia content through detailed, standardized descriptions. This standardization supports interoperability across different systems and platforms, allowing for seamless sharing and integration of multimedia information. Additionally, MPEG-7 aims to improve the precision and relevance of multimedia search and retrieval by providing rich metadata that can capture both the physical attributes and semantic content of multimedia resources. Through these capabilities, the MPEG-7 Ontology plays a critical role in advancing the field of multimedia information systems, driving innovation in how multimedia content is described, managed, and utilized.
- Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [1][11]

The Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) is a basic ontology intended to capture the underlying categories and connections that drive human cognition and linguistic structures. DOLCE, as part of the WonderWeb Foundational Ontologies Library, focuses on ontological clarity and semantic richness, offering a strong framework for the thorough and cohesive representation of knowledge across disciplines. DOLCE's core entities are endurant (objects that persist over time, such as physical objects and agents), perdurants (events and processes that occur over time, such as actions and occurrences), qualities (attributes and properties of objects and events, such as colour, shape, and duration), and abstract entities (non-physical concepts such as mathematical objects and propositions). These categories are meticulously defined to reflect the nuanced distinctions in how we perceive and describe the world, thus supporting the precise modeling of complex knowledge.

DOLCE covers a wide range of areas, including artificial intelligence, semantic web technologies, knowledge engineering, and cognitive science. Its principal goals are to improve the interoperability and reusability of ontologies by providing a common semantic framework that can be coupled with domain-specific ones. This integration facilitates the creation of semantically and logically compatible knowledge representations. DOLCE seeks to bridge the gap between human conceptualizations and formal ontological representations by emphasizing cognitively and linguistically meaningful concepts. This emphasis allows for more natural and effective information processing, as well as improved communication and cooperation between various systems and applications. DOLCE, as a fundamental ontology, not only helps to standardize ontological structures but also improves the accuracy and depth of semantic analysis in a variety of multidisciplinary fields.

- Core Ontology for Multimedia Annotation (COMM) [1][8]

  The Core Ontology for Multimedia Annotation (COMM) is a robust framework designed to facilitate the annotation and retrieval of multimedia content and is based on the MPEG-7 and DOLCE ontologies. COMM provides a structured and standardized approach to describing various multimedia elements, such as images, videos, audio files, and their associated metadata. By offering a common vocabulary and a set of well-defined relationships, COMM enables consistent and precise annotation, which is crucial for effective information retrieval, content management, and interoperability across different systems. Its comprehensive schema includes entities such as media assets, segments, annotations, and concepts, each of which plays a specific role in detailing the characteristics and contextual information of multimedia resources.

  The scope of COMM extends to various multimedia applications, including digital libraries, multimedia databases, content management systems, and semantic web technologies. Its primary objectives are to enhance the accessibility and discoverability of multimedia content by providing detailed and semantically rich annotations. COMM aims to bridge the gap between multimedia content and user queries by enabling sophisticated search mechanisms based on annotated metadata. Additionally, it supports the integration and interoperability of multimedia content from diverse sources, fostering a more interconnected and reusable digital ecosystem. By leveraging COMM, organizations can achieve better content organization, improved user experiences, and more effective utilization of multimedia resources.

The available **production & film related ontologies** in literature are the following:

- MovieLabs ontology

  The MovieLabs Ontology is a comprehensive framework developed by MovieLabs, a technology joint venture of major Hollywood studios, to standardize the representation and management of movie-related data across the industry. The ontology encompasses a wide array of entities crucial to the film ecosystem, including metadata for movies, TV shows, and other media assets. Key entities within

Funded by
the European Union

the ontology include titles, genres, cast and crew members, production and release details, rights information, and technical specifications of the media content. By providing a detailed and structured schema, the MovieLabs Ontology enables the seamless exchange and integration of data across different platforms and systems, fostering greater interoperability and efficiency in the media supply chain.

The scope of the MovieLabs Ontology extends beyond simple metadata representation to include the entire lifecycle of media content, from production through distribution to consumption. Its primary objectives are to enhance the accuracy and consistency of media data, streamline workflows, and support new business models and technologies in the entertainment industry. By standardizing data formats and definitions, the ontology aims to reduce friction in the exchange of information, enabling more effective collaboration between content creators, distributors, and service providers. Furthermore, the MovieLabs Ontology supports advanced applications such as personalized content recommendations, automated content processing, and enhanced search and discovery features, ultimately contributing to a more dynamic and responsive media ecosystem.

- OntoFilm ontology

  The OntoFilm ontology is a specialized framework designed to model and manage knowledge within the film industry, providing a detailed and structured representation of the various entities and relationships involved in film production, distribution, and consumption. The ontology includes entities such as films, directors, actors, roles, scenes, locations, production companies, and film festivals, among others. Each entity is meticulously defined, with properties and relationships that capture the complexity and interconnectivity of the film industry. For example, the ontology can represent the relationship between an actor and the roles they have played, or between a film and its shooting locations, facilitating a comprehensive and nuanced understanding of film-related data.

  The scope of the OntoFilm ontology encompasses the entire lifecycle of a film, from its inception and production to its release and reception by audiences and critics. Its primary objectives are to improve the organization, retrieval, and analysis of film data, supporting a wide range of applications from academic research and film archiving to recommendation systems and market analysis. By providing a common semantic framework, OntoFilm aims to enhance interoperability between different data sources and systems within the film industry, enabling more effective data integration and knowledge sharing. Furthermore, the ontology supports advanced querying and reasoning capabilities, allowing users to uncover insights and patterns in film data that would be difficult to detect using traditional data management approaches. This makes OntoFilm a valuable tool for filmmakers, researchers, and industry professionals looking to leverage structured knowledge to inform their work and decision-making processes.

- Yago-2 ontology

  YAGO-2 (Yet Another Great Ontology) is a large-scale, high-quality knowledge base that integrates and enhances information from Wikipedia, WordNet, and GeoNames. It is designed to provide a rich semantic network that connects entities and facts, forming a comprehensive ontology. YAGO-2 includes a wide range of entities such as people, organizations, locations, and events, each linked by well-defined relationships. For example, it can represent the birthplace of a person, the headquarters of an organization, or the geographical coordinates of a city. The ontology also captures temporal information, allowing it to represent events and the time periods during which they occurred, adding a valuable dimension to the knowledge it encompasses.

  The scope of YAGO-2 extends to various domains, making it a versatile resource for a multitude of applications, from semantic search and natural language processing to data integration and AI research. Its primary objectives are to ensure accuracy and comprehensiveness, leveraging the structured data from its sources to create a reliable and detailed ontology. By combining information from multiple datasets, YAGO-2 aims to provide a unified view of knowledge that is both broad and deep. Furthermore, YAGO-2 is designed to be highly scalable, supporting continuous updates and expansions to incorporate new data and maintain relevance. Its rigorous methodological framework and extensive coverage make YAGO-2 an invaluable tool for researchers, developers, and organizations seeking to leverage structured knowledge for advanced information retrieval and reasoning tasks.

3.2 Media Asset Manager

This section describes the current market solutions for media asset management, licensing, and distribution, analysing their advantages and disadvantages and comparing them with the Media Asset Manager and Distribution Engine tools offered by SCENE.

3.2.1 Bynder

Bynder[11] is a prominent Digital Asset Management (DAM) platform designed to help organizations streamline the organization, management, and distribution of digital assets. Known for its user-friendly interface, Bynder caters to businesses of all sizes, enabling them to maintain brand consistency and improve collaboration across teams.

Bynder offers a range of robust features, including intuitive metadata tagging for easy asset retrieval, comprehensive version control to track asset changes, and advanced search functionalities to quickly locate files. Additionally, Bynder supports customizable dashboards, approval workflows, and integration with popular third-party applications, making it a versatile tool for managing digital content.

Despite its strengths, Bynder has some limitations. It offers limited capabilities for managing licenses and rights, which may necessitate additional tools for comprehensive licensing management. The platform can also be cost-prohibitive for smaller businesses, and some advanced features may require additional training to utilize effectively. Integration with other systems, while available, can sometimes be complex to set up and maintain.

3.2.2 Adobe Experience Manager Assets

Adobe Experience Manager (AEM) Assets[12] is a comprehensive Digital Asset Management (DAM) solution integrated within the Adobe Experience Cloud. Designed for large enterprises, AEM Assets provides extensive capabilities for managing, storing, and distributing digital content, offering seamless integration with Adobe's suite of creative tools.

AEM Assets boasts a wide array of features, including advanced version control to manage asset iterations, detailed metadata management for efficient organization, and powerful search functionalities to locate assets quickly. It supports robust integration with Adobe Creative Cloud, facilitating a streamlined workflow for creative and marketing teams. Additionally, AEM Assets offers customizable approval workflows, scalability to handle large volumes of assets, and tools for maintaining brand consistency.

While AEM Assets is a powerful DAM solution, it comes with certain limitations. The platform can be expensive, making it less accessible for small to medium-sized businesses. Its complexity may require significant training and resources to implement and manage effectively. Additionally, the extensive feature set can be overwhelming for new users, potentially leading to a steeper learning curve.

---

[11] https://www.bynder.com/en/ (Accessed on 26 July 2024).

[12] https://business.adobe.com/products/experience-manager/adobe-experience-manager.html. (Accessed on 26 July 2024).

Funded by
the European Union

3.2.3 Canto

Canto[13] is a Digital Asset Management (DAM) platform tailored for small to medium-sized businesses. Known for its ease of use and affordability, Canto helps organizations efficiently organize, manage, and distribute their digital assets, ensuring streamlined workflows and improved collaboration.

Canto offers a variety of user-friendly features, including strong organizational capabilities with detailed metadata tagging and an intuitive search function to quickly locate assets. The platform provides customizable workflows and version control to manage asset updates and approvals. It also supports easy sharing and collaboration through secure, branded portals, making it simple for teams to access and utilize digital content.

Despite its strengths, Canto has some limitations. It offers limited licensing and rights management capabilities, which may require additional tools for comprehensive license tracking. The platform, while affordable, may lack some of the advanced features needed by larger enterprises or more complex workflows. Additionally, while Canto is easy to use, the customization options can be less extensive compared to more robust DAM systems, potentially limiting its adaptability for highly specialized needs.

3.2.4 Widen Collective

Widen Collective[14] is a comprehensive Digital Asset Management (DAM) platform designed to meet the needs of organizations of all sizes. It focuses on providing robust tools for managing, organizing, and distributing digital assets, with an emphasis on content governance and analytics to enhance overall asset management efficiency.

Widen Collective offers a wide range of features, including detailed metadata tagging and advanced search capabilities for easy asset retrieval. It provides version control to manage asset updates and ensures content consistency. The platform also includes customizable workflows for content approvals and robust analytics to track asset usage and performance. Widen Collective is highly customizable, allowing organizations to tailor the platform to their specific needs, and it integrates with various third-party applications to support seamless workflows.

While Widen Collective is a powerful DAM solution, it has some limitations. The platform can be less intuitive for new users, requiring a learning curve to fully utilize its extensive features. The high level of customization, while beneficial, can add to the complexity of setup and management. Additionally, the comprehensive feature set and customization options can make it a more expensive solution, potentially limiting its accessibility for smaller businesses or those with limited budgets.

3.2.5 Market analysis and comparison with SCENE

Table 1 provides a comprehensive comparison of various tools used in media licensing and distribution, including Bynder, Adobe Experience Manager (AEM) Assets, Canto, and Widen Collective, together with the SCENE platform to be developed during the project. Each tool is evaluated across several key functionalities: user interface, scalability, license management, media distribution, decentralized license agreements, ease of implementation, and customizable distribution platform. This comparison aims to highlight the strengths and limitations of each tool, offering insights into their suitability for different business needs and technical environments. This table illustrates how SCENE's Media Asset Manager and Distribution Engine can uniquely position themselves in the market by offering advanced features that address the limitations of current solutions.

[13] https://www.canto.com/. (Accessed on 26 July 2024).
[14] https://www.acquia.com/products/acquia-dam. (Accessed on 26 July 2024).

Table 1: Comparison between current market solutions and the SCENE platform

Platform	User Interface	Scalability	License Management	Media Distribution	Decentralized License Agreements	Ease of implementation	Customizable Distribution platform
Bynder	Intuitive and user-friendly	Scalable for various business sizes	Basic	Basic	No	High	No
Adobe Experience Manager Assets	Integrated with Adobe suite, complex	Suitable only for large enterprises	Basic	Requires Adobe eco-system	No	Low	Yes
Canto	Intuitive and user-friendly	Suitable only for small to medium businesses	Basic	Basic	No	High	Limited
Widen Collective	Less intuitive for new users	Suitable for all sizes	Comprehensive	Advanced	No	Medium	Yes
SCENE	Intuitive and user-friendly	Suitable for all sizes	Comprehensive	Advanced	Yes	High	Yes

3.3 3D Model Reconstruction

The primary goal of 3D model reconstruction tool is to generate detailed and accurate 3D representations of cultural heritage sites. This involves the utilization of advanced hardware and software tools to capture and process data using state-of-the-art 3D modelling tools to create comprehensive 3D models that reflect the intricate details of these sites. The aim is to integrate these 3D models into virtual reality (VR) platforms, allowing for immersive and interactive exploration. This will enhance the accessibility and engagement with cultural heritage sites, providing users with a rich, virtual experience.

3.3.1 3D Reconstruction Pipeline

For the development and implementation the 3D model reconstruction tool, we have designed and developed an architecture diagram, depicted in Figure 3 which showcases the components/processes within the 3D model reconstruction tool. In figure 2.3.1, we can see that after a location of interest is scanned, the source point clouds and imagery are fed into this pipeline as inputs. After this, the inputs are used for two purposes – for storing the resources in the *Data Lake* for future use and for processing the inputs for creating/rendering 3D models.

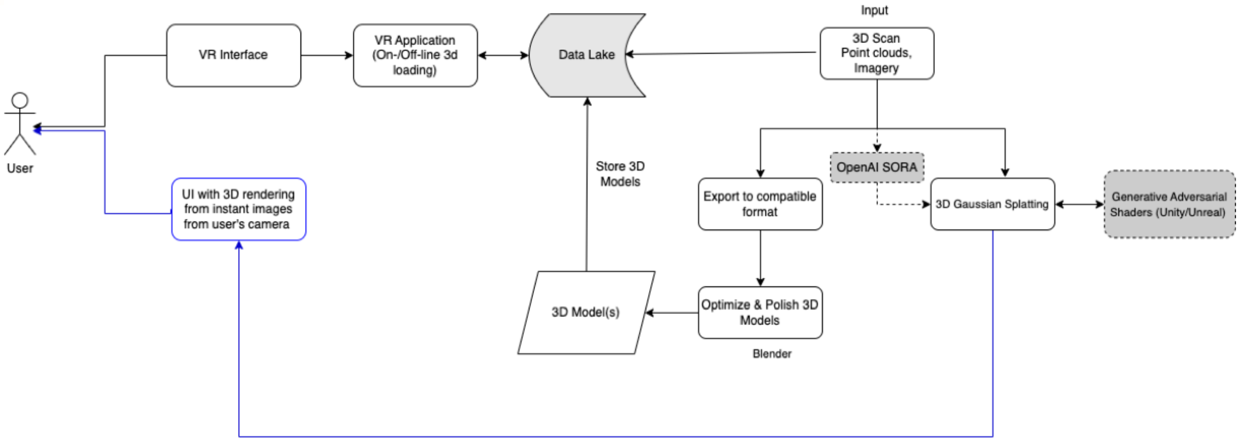


Figure 3: Architecture diagram for 3D model reconstruction of cultural heritage sites.

For the creation/rendering of the 3D model, the inputs are passed to either of the two modules – Export to compatible format and automatic 3D modelling tools like 3D Gaussian Splatting or NeRF; there is a 3rd module, OpenAI SORA, marked with dotted rectangle, which is in the planning phase and it is about incorporating Generative-AI into the automatic 3D model reconstruction pipeline. The module Export to compatible format converts the input point clouds/imagery into a suitable format for enhancing the quality of the 3D model using state-of-the-art tools like Blender. The enhanced 3D model is then exported into suitable format, e.g., .fbx and finally stored in the Data Lake for further use, e.g. In the VR environments. The other module, automatic 3D modelling tools like 3D Gaussian Splatting or NeRF, takes the inputs and renders 3D view for the captured scene. Here the existing or novel views are presented to the users in an interactive UI where the users can view the 3D scene from various viewing angles. Using this module, the users (e.g., location scouter, directors) would be able to capture their own set of scenes and use these tools to create a rather spontaneous/quick 3D rendering of their own scenes. For this component, we also plan to incorporate Generative-AI tools such as OpenAI SORA in the future for further enhancement in the automatic 3D rendering of captured scenes. Furthermore, we plan to investigate Generative Adversarial Shaders (in Unity/Unreal Engine), marked with dotted rectangle in Figure 3, for optimizing the total processing pipeline. In the next, a detailed descriptions of the technologies involved in this pipeline is presented.

3.3.2 Brief Summary of Technologies Used for 3D Model Reconstruction

For 3D model reconstruction, we used state-of-the-art automatic tools such as NeRF, 3D Gaussian splatting, and Photogrammetry. Moreover, we also used *Blender* as part of enhancing the 3D models manually. Below we present a summary of the used technologies and the results which we achieved as of now.

3.3.2.1 3D model using NeRF and 3D Gaussian Splatting

For the 3D model reconstruction, we implemented, among others, the Neural Radiance Fields (NeRF) models and 3D Gaussian Splatting. We have successfully trained the NeRF model using the scans and imagery provided by the project partners GOF. This training is crucial for achieving high-quality 3D reconstructions. After the training was completed, we are able to view the 3D scene from existing and novel views. Figure 4 shows a rendered view of a Pilot site "Holy Church of Panagia of Asinou, Cyprus"; here we can see 2 different render modes - *A0* and *shades* which show different mode of visualisation of the rendered 3D scene. To handle larger datasets, we have explored the ZIP-NeRF method, which facilitates the loading and processing of extensive point clouds, enhancing the scalability of the 3D models. Figure 5 shows another 3D model using these methods.

Render mode: AO *Render mode: Shade*

Figure 4: 3D model reconstruction of a Pilot site with NeRF.

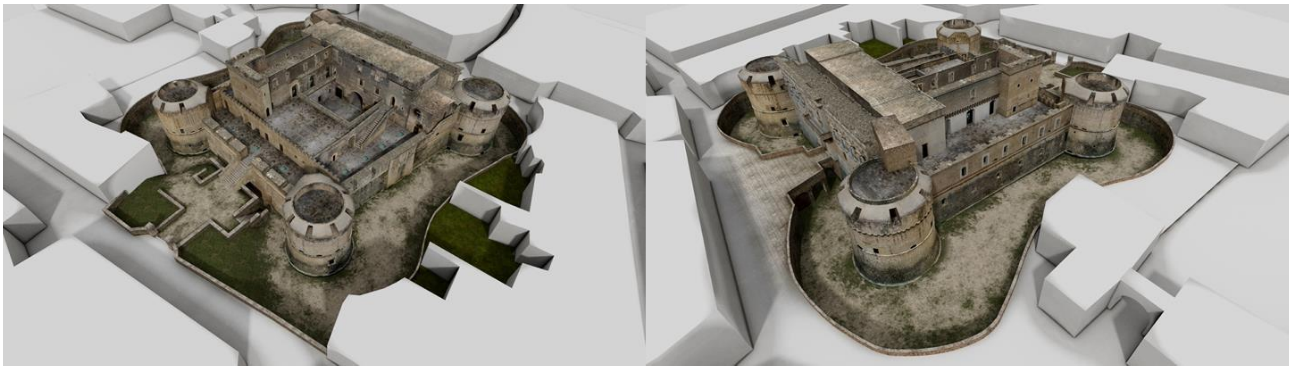
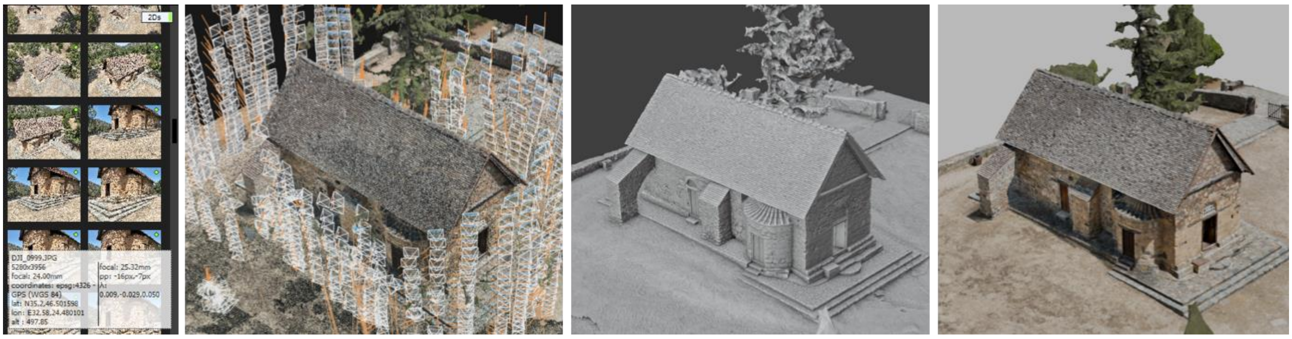


Figure 5: Highly optimized 3D model of a Pilot site (Castel de Monti-Corigliano d'Otranto (LE)) with optimized topology for performance reference and final use.

3.3.2.2 3D model using Photogrammetry and Blender

Along with NeRF and 3D Gaussian Splatting, we also used state-of-the-art photogrammetry tools (e.g., Reality Capture) and Blender for reconstructing high-quality 3D models from the scanned imagery.

Photogrammetry uses a series of images (depending on the size of the location and geometry it could be 1000 to 8000 images per site); this technique involves capturing multiple photographs of an object or site from different angles and processing them to generate accurate 3D representations. It uses camera alignment parameters and nearly a million trigons (triangular polygon) to create a 3D model which is finally textured with high-resolution image textures. Figure 6 depicts 3D models of 2 pilot sites.



Funded by the European Union

a. Captured photos *b. Camera alignment* *c. 3D model* *d. Textured model*

Figure 6: 3D model generation of Pilot sites with Photogrammetry.

We also used Blender to create 3D models from the scanned point clouds and imagery. In Blender the point cloud needs to be converted to a mesh using a modifier (Remesh modifier or external software), and refine the mesh with Blender's editing tools. Then, the mesh needs to be unwrapped to create a UV map and then textures are applied using the captured images. Finally, lighting and rendering are setup and then the model is exported in the desired format. Figure 7 shows a 3D model reconstructed in Blender for a pilot site.

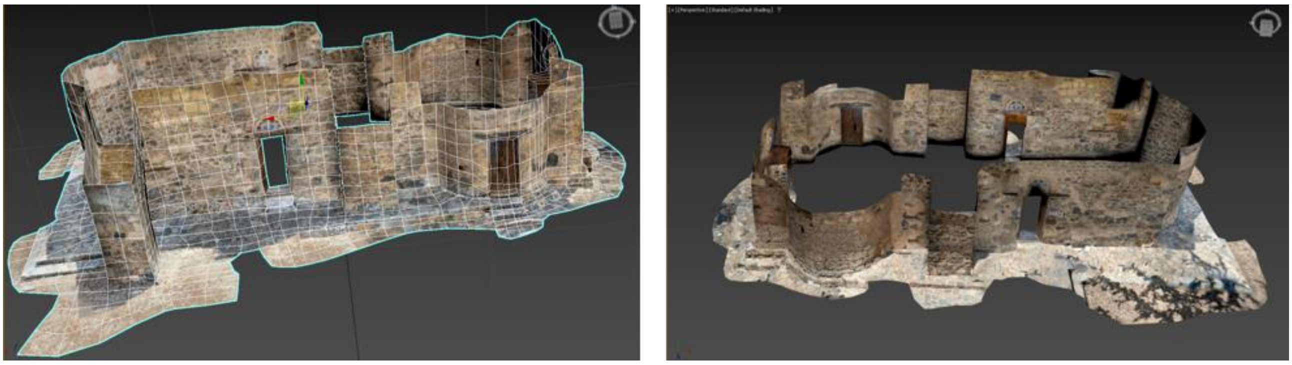


Figure 7: 3D model of a Pilot site (Holy Church of Panagia of Asinou, Cyprus) with Blender.

3.3.3 State-of-the-art of the Technologies used for 3D Model Reconstruction

In 2.3.3, we have briefly described different tools and technologies which are being used by the SCENE projects partners to create/reconstruct high-quality 3D models from the captured imagery and point clouds. Below, we present an analysis of the state-of-the art technologies including the scene/location capturing technology and 3D rendering/reconstruction tools/technology in this sector.

3.3.4 State-of-the-art technology for capturing scene from different viewing angles

Laser scanning: Laser scanning is a highly precise method for creating 3D models from a series of laser beams that measure the distance and shape of an object or surface. This technology is capable of capturing high-resolution details and accurate measurements of complex geometries, such as statues or architectural elements. The process involves:

- Laser Scanner: A device that emits laser beams and measures the time it takes for the beams to return after hitting an object. Laser scanners can be handheld or mounted on a tripod. Examples include the Leica Cyclone[15], FARO Focus 3D[16], and Trimble TX8.
- Computer: Used to process the data collected by the laser scanner.

[15] https://leica-geosystems.com/products/laser-scanners/software (Accessed on 26 July 2024).
[16] https://www.faro.com/de-DE/Products/Software/SCENE-Software (Accessed on 26 July 2024).

- Software: Specialized software is needed to process the point cloud data and create a 3D mesh. Some examples of laser scanning software are:
  - Leica Cyclone: Offers robust point cloud processing capabilities and is widely used in various industries for detailed 3D documentation.
  - FARO Scene: Provides intuitive workflows for the automatic processing and analysis of 3D laser scan data.
  - CloudCompare[17]: An open-source software for processing 3D point clouds and meshes, known for its versatility and extensive plug-in support.
  - Geomagic[18]: A suite of 3D software tools for reverse engineering, 3D inspection, and CAD modeling.

Multi-sensor approaches: Multi-sensor approaches involve the integration of different types of sensors to create comprehensive 3D models that capture various aspects of an object or site. This method enhances the quality and completeness of 3D models and can reveal hidden features or information not visible to the naked eye. The process includes:

- Sensors: Multiple devices such as cameras, laser scanners, infrared sensors, radar sensors, and more. Each sensor type provides unique data that, when combined, offer a more complete picture.
- Computer: Required for processing and fusing the data collected from the different sensors.
- Software: Specialized software is used to fuse data from different sources and create a 3D mesh. Examples of multi-sensor software include:
  - Autodesk Recap[19]: Facilitates reality capture and integrates seamlessly with Autodesk's suite of design tools, enabling the creation of detailed 3D models from various data sources.
  - Bentley ContextCapture[20]: Allows for the production of high-resolution 3D models from simple photographs and/or laser scans, offering powerful photogrammetric capabilities.
  - Pix4Dmapper[21]: A photogrammetry software for professional drone mapping, turning images into highly accurate 2D maps and 3D models.
  - Eos Systems PhotoModeler[22]: Specializes in extracting accurate and detailed 3D measurements and models from photographs.

Structured Light Scanning: Structured Light Scanning involves projecting a series of light patterns onto an object and capturing the deformation of these patterns with cameras. This method is known for its high accuracy and speed in capturing 3D data. The process includes:

- Projectors and Cameras: A structured light scanner typically includes one or more projectors and cameras.
- Computer: Used to analyze the captured images and reconstruct the 3D model.
- Software: Processes the deformation of the light patterns to generate a detailed 3D model. Examples include:
  - Artec Studio[23]: A professional-grade software that provides tools for 3D scanning and data processing, offering powerful algorithms for high-accuracy 3D modeling.

[17] https://github.com/CloudCompare/CloudCompare (Accessed on 26 July 2024).
[18] https://www.3dsystems.com/software/geomagic-design-x (Accessed on 26 July 2024).
[19] https://www.autodesk.de/products/recap/overview?term=1-YEAR&tab=subscription (Accessed on 26 July 2024).
[20] https://working-system.de/produkte/contextcapture/ (Accessed on 26 July 2024).
[21] https://www.pix4d.com/de/produkt/pix4dmapper-photogrammetrie-software/ (Accessed on 26 July 2024).
[22] https://www.photomodeler.com/ (Accessed on 26 July 2024).
[23] https://www.artec3d.com/3d-software/artec-studio (Accessed on 26 July 2024).

Funded by the European Union

- GOM Inspect [24]: Comprehensive 3D inspection and mesh processing software for analyzing 3D point clouds from structured light scanners.

3.3.5 State-of-the-art technology for 3D modeling using captured point clouds and imagery

Here we present an analysis of the state-of-the-art technologies for 3D modeling using data from captured point clouds and imagery.

Photogrammetry: Photogrammetry is an essential technology for capturing scenes from different viewing angles, utilizing photographs to create 3D models. The process involves:

- Cameras: High-resolution digital cameras are used to capture multiple overlapping photos of an object or scene from different angles.
- Computer: Used to process and stitch together the captured images.
- Software: Photogrammetry software analyzes the photographs to create 3D models. Examples include:
    - Agisoft Metashape[25]: Provides professional photogrammetric processing of digital images and generates 3D spatial data to be used in GIS applications, cultural heritage documentation, and visual effects production.
    - RealityCapture[26]: A state-of-the-art photogrammetry software that creates virtual reality scenes, textured 3D meshes, orthographic projections, geo-referenced maps, and more from images or laser scans.
    - Meshroom[27]: Meshroom is an open-source 3D Reconstruction Software based on the AliceVision framework which is a Photogrammetric Computer Vision Framework which provides 3D Reconstruction and Camera Tracking algorithms.
    - 3DF Zephyr[28]: Allows the reconstruction of 3D models from photos with a user-friendly interface and powerful processing capabilities.

Structure from Motion (SfM): SfM [1][58][1][59] is a photogrammetric technique that creates 3D structures from 2D image sequences. Unlike traditional photogrammetry, SfM automatically matches features across images to build 3D models. The process includes:

- Cameras: Multiple images taken from different angles.
- Computer: To process and stitch the images together.
- Software: SfM software identifies and matches features across images to reconstruct the 3D structure. Examples include:
    - VisualSfM [1][58]: A 3D reconstruction software for visualizing large-scale scenes.
    - COLMAP [1][59]: A photogrammetry pipeline that uses SfM for 3D model reconstruction.

Volumetric Video: Volumetric Video [1][58] captures 3D data in real-time to create dynamic, realistic 3D models of moving subjects. This technology is often used in virtual reality (VR) and augmented reality (AR) applications. The process includes:

- Multiple Cameras: A rig of cameras placed around the subject to capture different angles simultaneously.
- Computer: High-performance computing for real-time processing.
- Software: Platforms that stitch and process the video feeds into a coherent 3D model. Examples include:

[24] https://www.gom.com/en/products/software/gom-inspect (Accessed on 26 July 2024).
[25] https://www.agisoft.com/ (Accessed on 26 July 2024).
[26] https://www.capturingreality.com/ (Accessed on 26 July 2024).
[27] https://meshroom-manual.readthedocs.io/en/latest/ (Accessed on 26 July 2024).
[28] https://www.3dflow.net/3df-zephyr-photogrammetry-software/ (Accessed on 26 July 2024).

Funded by the European Union

- Microsoft Mixed Reality Capture Studio[29]: Provides tools and technology for capturing and processing volumetric video.
- 8i[30]: A company that offers software for creating holographic content from volumetric video.

Light Field Capture: Light Field Capture records the amount of light traveling in every direction through every point in space, providing extensive information about the scene. This method allows for post-capture focusing, depth estimation, and perspective shifts. The process involves:

- Light Field Cameras: Specialized cameras, like the Lytro camera, that capture light fields.
- Computer: For processing and rendering [1][60] the captured light field data.
- Software: Tools to process light fields and generate 3D models. Examples include:
  - Lytro Desktop: Software for processing light field data captured by Lytro cameras.
  - Adobe Photoshop[31][23]: Has features for processing light field data to a limited extent.

Neural Radiance Fields (NeRF): NeRF is an innovative technology that uses deep learning to generate high-quality 3D representations from 2D images. NeRF models learn to encode the volumetric scene function, allowing for photo-realistic rendering of complex scenes from novel viewpoints. The process includes:

- NeRF Models: Neural networks that are trained using a set of 2D images taken from different angles to learn the underlying 3D structure of the scene.
- Computer: Used to train the neural network and process the data.
- Software: Frameworks and libraries for training NeRF models and generating 3D scenes. Examples include:
  - NeRF [61] (original implementation): Developed by researchers at UC Berkeley and Google Research, the original NeRF implementation leverages deep learning for novel view synthesis.
  - PlenOctrees [62]: A technique that accelerates NeRF rendering by combining neural networks with efficient spatial data structures.
  - Mip-NeRF [63]: An extension of NeRF that improves performance and quality by using multiscale input representations.
  - Zip-NeRF [64]: An extension of NeRF that handles bigger models and improves the ovverall surface quality and processing speed.

3D Gaussian Splatting: 3D Gaussian Splatting is a technique used to render point clouds with smooth surfaces by representing each point as a 3D Gaussian distribution. This method helps in creating high-quality visualizations of 3D data and can be combined with other techniques for improved results. The process includes:

- 3D Gaussian Splatting Algorithm [65] [66]: Processes point cloud data to generate Gaussian splats that represent the surface of the object.
- Computer: Used to process and visualize the data.
- Software: Tools and libraries for implementing 3D Gaussian splatting. Examples include:
  - Point Cloud Library (PCL) [67]: An open-source library for 3D point cloud processing that supports various techniques including Gaussian splatting.
  - Open3D [68]: Another open-source library designed for 3D data processing, visualization, and analysis, which can be used to implement Gaussian splatting techniques.

29 https://transforminteractive.com/mixed-reality-capture-studios-a-look-inside/ (Accessed on 26 July 2024).
30 https://8i.com/ (Accessed on 26 July 2024).
31 https://www.adobe.com/ (Accessed on 26 July 2024).

3.4 Blockchain-based IPR preservation

The Blockchain framework aims advance Intellectual Property Rights (IPR) management by leveraging cutting-edge technologies. The technical analysis of the technologies employed, emphasizing their specific implementations within SCENE and the innovative advancements they bring to the field, are presented here.

3.4.1 Current State-of-the-Art:

Existing blockchain-based IPR management systems predominantly utilize:

- **Smart Contracts:** These self-executing agreements, often written in domain-specific languages or general-purpose programming languages like Solidity, automate processes like licensing and royalty distribution. However, they often lack the legal rigor and human readability required for complex IPR agreements, which SCENE addresses through the integration of Ricardian Contracts.
- **Blockchain Architecture:** Decentralized ledgers, such as Ethereum or Hyperledger Fabric, ensure secure and transparent recording of transactions and ownership information. However, scalability and privacy concerns remain challenges for large-scale IPR management. SCENE leverages the permissioned nature of Hyperledger Fabric to address these concerns, providing fine-grained access control and privacy features tailored to the specific needs of IPR management.
- **Off-Chain Storage and Content Signatures:** To overcome blockchain storage limitations, large files like media assets are typically stored off-chain, with their cryptographic hashes recorded on the blockchain. Content signatures are used to verify the integrity and authenticity of these assets. SCENE utilizes IPFS for decentralized storage and integrates content signatures into its NFT and Ricardian Contract for enhanced security and traceability.

3.4.2 SCENE's Technological Innovation:

SCENE enhances current practices through the following technical implementations:

- **Custom Golang Chaincodes (Smart Contracts) for NFTs and Ricardian Contracts:**
  - NFT Chaincode Implementation: SCENE's custom Golang chaincode for NFTs defines a comprehensive data structure within Hyperledger Fabric to represent NFT metadata, including ownership details, licensing terms, associated media files (referenced via IPFS hashes), and transaction history. The chaincode implements functions for NFT creation (minting), ownership transfer, and royalty management, ensuring secure and transparent transactions while maintaining compliance with Ricardian Contract terms.
  - Ricardian Contract Chaincode Implementation: The Ricardian Contract chaincode in SCENE leverages Golang's powerful string manipulation and parsing capabilities to extract structured data from the contract's natural language text. This structured data is then stored alongside the original text, enabling automated verification and enforcement of contract terms. The chaincode also includes functions for dispute resolution, leveraging the decentralized nature of the blockchain for impartial decision-making.
- **Secure Media Distribution and Tracking:**
  - HLF Gateway API Service Implementation: SCENE's custom HLF Gateway API service acts as a bridge between external applications (e.g., media distribution platforms) and the Hyperledger Fabric blockchain. It exposes RESTful endpoints for interacting with the NFT and Ricardian Contract chaincodes, handling authentication, authorization, and data validation to ensure secure and controlled access to the blockchain's functionality.
  - IPFS Content Addressing and Integration: SCENE utilizes IPFS's content addressing scheme to assign unique identifiers (hashes) to media files. These hashes are stored in the NFT metadata on the blockchain, providing a decentralized and tamper-proof way to reference

and retrieve the associated media assets. IPFS's peer-to-peer file sharing protocol enables efficient and resilient content distribution.

- Blockchain-Based Transaction Logging Implementation: Every transaction involving media assets, such as licensing agreements or ownership transfers, is recorded as a transaction on the Hyperledger Fabric blockchain. This transaction data includes timestamps, involved parties, and references to the relevant NFTs and Ricardian Contracts. The immutable nature of the blockchain ensures a transparent and auditable record of all IPR-related activities.

3.4.3 SCENE's Technological Stack:

The main technologies which participate in the development and deployment of SCENE Blockchain framework are:

- **Hyperledger Fabric:** A permissioned blockchain platform chosen for its flexibility, scalability, and privacy features. Fabric's channel architecture enables the creation of separate communication channels for different stakeholders, ensuring data confidentiality and compliance with regulatory requirements.
- **IPFS (InterPlanetary File System):** A decentralized file system that complements Hyperledger Fabric's capabilities by providing a distributed and fault-tolerant storage solution for media assets. IPFS's content addressing ensures data integrity and eliminates the risk of single points of failure.
- **Golang:** The Go programming language, selected for its efficiency, concurrency support, and strong community backing in the blockchain development space. Golang's static typing and garbage collection features contribute to the development of robust and secure chaincodes.
- **Docker:** Containerization technology used to encapsulate the SCENE components into isolated environments, ensuring consistent deployment and portability across different systems.
- **HLF Gateway:** A custom-developed API gateway that simplifies interaction with the Hyperledger Fabric network. It provides a user-friendly interface for external applications to access and manipulate chaincode data, abstracting the complexities of blockchain communication.

3.5 Location Scouting

The Location Scouting tool, has been designed to facilitate the process of finding the most suitable filming locations for filmmakers. This tool leverages state of the art technologies to register, index, and search multimedia content available for all locations, effectively cataloguing points of interest (POIs) that are ideal for various shooting needs, while extracting automatically metadata via image captioning methods. By facilitating the search of the most suitable POIs based on specific search criteria, the tool ensures that film producers can effortlessly discover and select optimal locations for their scenes and save time, and enhance their creative vision. In addition, this tool is enhanced with a chatbot application, allowing its end-users to search for the locations' information in a user-friendly way.

3.5.1 Image captioning

Image captioning is a multidisciplinary activity that combines computer vision and natural language processing with the goal of producing descriptive textual captions for provided pictures. Recent advances in deep learning, notably the use of convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have considerably improved picture captioning systems' performance. This review delves into the most advanced approaches, systems, and issues in picture captioning today. [38]

The encoder-decoder architecture is the most often used technique for picture captioning. To extract rich feature representations from the input picture, a CNN, often a pre-trained model such as ResNet [39] or Inception, serves as the encoder. These characteristics are then fed into an RNN, often a Long Short-Term

Memory (LSTM) network or a Gated Recurrent Unit (GRU), which serves as the decoder and generates the accompanying textual description. The incorporation of attention methods, presented by Xu et al. in the "Show, Attend, and Tell" model [40], has boosted performance even further by allowing the decoder to dynamically focus on different regions of the image while producing each word in the caption.

Transformer-based approaches, such as the Vision Transformer (ViT) [41] and Vision-and-Language Pre-training (VLP) frameworks [42], have lately proven to be more effective in picture captioning. These models use the self-attention process to identify global dependencies and connections inside and between images and text. Vision-language models, such as ViLBERT [43] and CLIP [44], have been pre-trained on huge datasets to learn combined representations of visual and textual information, resulting in considerable increases in caption generation accuracy and contextual relevance.

3.5.2 Chatbot

Advances in AI and NLP have facilitated the development of chatbot technology in a variety of sectors. This section examines relevant research that serves as a foundation for the creation, application, and assessment of chatbots in public administration. According to Androutsopoulou et al. [46], chatbots should become one of the main communication channels between citizens and government. A recent case study done by [46] intended to identify various stakeholders' issues, as well as the constraints and facilitators that impact the design of chatbots in the public sector in Ukraine. According to the study's conclusions, based on a careful assessment of the state of the art and twelve interviews with specialists participating in the LvivCityHelper[32] bot project, the first step before the implementation and adoption of a chatbot in public services concerns the identification of the factors affecting the introduction of chatbots in the society.

Nirala et al. [47] address AI-based chatbots, covering applications, issues, architecture, and models. They note the progression of chatbots, from Turing Test and Rule-based chatbots to sophisticated Artificial Intelligence-based chatbots. Their poll sought to better understand and examine the possibilities of chatbots in consumer and public administration services. The survey's findings revealed that AI-assisted chatbot systems have significant potential for boosting customer service and governance in public administration services.

There are three basic kinds of chatbots: template-based, corpus-based, intent and rule-based, AI-based, and hybrid. These techniques are available to handle different degrees of complexity, customization, and functionality in chatbot development [48] [49]. The complexity of the use case, available resources, desired level of customization, and intended user experience all influence the implementation method that is chosen. In addition, the appropriate technique is chosen based on the nature of the conversational agent's task and the data used. Each strategy's attributes are summarized below. [50]

Corpus-based chatbots employ massive volumes of text data, or corpora, to analyse user queries and provide replies. By analysing patterns and context in the corpus, these chatbots can have a better knowledge of linguistic subtleties, context, and user intent, resulting in more accurate and relevant replies. They can also tailor their replies based on the context of the discussion and the information included in the corpus. This flexibility enables them to give more nuanced and contextually relevant solutions, hence improving the overall user experience and engagement. This is owing to their design, which allows them to be trained on corpora from many domains and themes, making them adaptable to a wide range of applications and industries. Corpus-based chatbots can be used for customer assistance, healthcare, banking, and education. [51] [52]

Intent-based chatbots excel at identifying the underlying intent or purpose of user messages. Natural language understanding (NLU) techniques enable accurate identification and response to user intentions, resulting in more relevant and helpful interactions [53] [54]. They are intended to give replies that are specific to the indicated user intent. This concentrated strategy guarantees that the chatbot responds to customer

[32] https://city-helper.com/ , (Accessed on 24 July 2024).

inquiries directly and efficiently, resulting in a more seamless user experience and improved satisfaction levels. One of their key advantages is that they follow an organised conversation flow based on specified intentions and answers. This organised method promotes better information and simpler navigation for users, decreasing confusion and increasing engagement.

Rule-based chatbots are relatively straightforward to design, implement, and maintain compared to AI-based approaches. They operate on predefined rules and decision trees, making them predictable and easier to troubleshoot. They also follow a predetermined conversational flow based on predefined rules and patterns. This controlled flow allows for consistent interactions and ensures that the chatbot stays within the intended scope of the conversation. One of their weaknesses is that they have limited flexibility and adaptability compared to AI-based approaches. Also, they heavily rely on predefined rules and patterns to interpret user inputs and generate responses. This dependency can result in limited coverage for handling unexpected or diverse user queries that fall outside the scope of the predefined rules. They struggle to handle complex queries, understand natural language expressions, and adapt to evolving user needs without constant updates to the rule set. As the complexity of conversational scenarios increases, the rule set of the chatbot may become cumbersome to manage and maintain. Adding new rules or modifying existing ones to accommodate changes in user behavior or conversation patterns can be time-consuming and labor-intensive. [55] [56]

AI-powered chatbots may learn and adapt to new data and user interactions over time. These chatbots can enhance their performance, understand user preferences, and give more personalized replies with the help of machine learning algorithms and continual training. They can handle sophisticated inquiries and discussions, including ones with numerous intentions or layers of context. They can negotiate confusing or complex linguistic patterns to extract relevant information and respond appropriately. However, AI-powered chatbots rely significantly on massive volumes of training data to efficiently understand and generalize trends. Inadequate or biased training data might result in poor performance or promote negative behaviors and reactions. Furthermore, the inner workings of AI-powered chatbots, particularly those built on deep learning models, may be complicated and opaque.

3.6 AI-based Audience Preference Scouting

Introducing the AI-based Audience Preference Scouting advanced toolkit, designed to predict trends by analyzing audience viewing behavior and their interactions with content. Powered AI, this toolkit is trained to recognize and forecast trends, providing valuable insights for content creators and marketers. To ensure viewer privacy, the system employs persona creation, which anonymizes individual data while preserving essential behavioral patterns. More specifically, sentiment detection is enhanced through OpenCV-based facial emotion recognition [57], allowing for a nuanced understanding of audience reactions. This comprehensive solution offers a powerful blend of privacy-conscious analytics and predictive capabilities, driving informed decision-making in content development and strategy.

State of the art of trend scouting depends on the position within the production process.

Preproduction phase

In the preproduction phase, trend scouting is generally based on two factors.

1. Performance of previous, similar productions
2. Trending topics

The first uses predictive analysis, utilizing statistical methods, often AI supported, based on box office performance, streaming numbers and overall financial performance of similar films.

This approach has a high predictive value if planned new productions are similar to existing ones. This is particularly true for sequels or instalments in franchises. The further a new production departs from previous productions, the less meaningful the trend forecast becomes.

For productions dissimilar to recent other productions, general trend forecasts can utilize, based on social media posts, general audience sentiment and long-period topical trends derived from search engines.

Funded by
the European Union

Production phase

During the production phase, the audience preference assessment shifts from the issue which movie should be made, which is the core question in pre-production, to the assessment on how to optimize the movie currently made to reach the widest possible audience for this movie.

This, generally, will involve a presentation of the movie to a limited test audience to measure the reaction. The primary method for this is the use of test screenings, where a preview audience, which is either selected to consist of a reasonably representative cross-section of the population or is representative of the assumed target audience, is shown a version of the finished film. Audience reactions are then either elicited via interviews or questionnaires, or test screenings are performed in purpose-built movie theatres, allowing direct feedback to scenes via a set of rating buttons.

In special cases, direct measurements of reactions of individual viewers are performed, utilizing, for example, eye tracking, reaction speed and behavior analysis. This, however, is primarily used in the domain of advertisement production (e.g., eye-square.com), as here often the reaction of viewers to every second of a short ad is important. Such detailed analysis is still rarely used in the context of movie production.

Post-Production phase

During the post production phase, the trend forecasting shifts from the movie being made to its optimum distribution. The core questions here are in which contexts the movie should be promoted, with what other productions it should be bundled or recommended together, on which channels it should be distributed, to what target audiences and at what dates and times.

Existing recommender systems are highly predictive regarding target audiences and presentation context. There are, however, few current systems supporting scheduling decisions, as sufficiently rich data sources in the area have only become recently available.

3.7 Audience Building

The aim of the Audience Building tool is to revolutionize audience engagement and crowdfunding support for new productions. This platform aims to harness the power of gamification techniques to captivate and retain audiences, making participation both fun and rewarding. By launching audience-building campaigns and employing advanced social media monitoring, the tool provides valuable insights into the current impact and reach of your production, while by the use of sentiment analysis tool it will be able to provide valuable information about the sentiment of the audience for a certain production. Additionally, it streamlines the process of securing funding by connecting with funding agencies and directly engaging with the audience. Leveraging the robust blockchain infrastructure developed in T3.5, the platform also enables the use of NFTs, allowing producers to offer unique digital assets to their supporters, further enhancing engagement and funding opportunities. In addition, the tool provides to the producer an additional functionality, the summarization of the collected feedback, facilitating thus the campaigns creation used for the increased audience engagement.

The following sub-sections present a literature review for the sentiment analysis and summarization AI models that will be implemented and supported by this tool.

3.7.1 Sentiment analysis

Sentiment categorization is separated into two categories: sentence-level classification and document-level classification. To determine if a document (for example, an entire online post) reflects a broad positive or negative viewpoint, document-level sentiment classification examines the overall direction and polarisation of sentiment within the content. Sentiment analysis is often viewed as a subset of document analysis. text representation is an important step in this categorization since it must correctly reflect the original information that words or phrases in a text were meant to convey.

The bag-of-words (BoW) concept has long been used to create text representations. A document is turned into a vector based on the BoW format that has a defined length and equals the vocabulary size. Each component of the representation vector references a word's presence in the text. Alternatively, a vector element might represent a word's frequency or TF-IDF score, depending on whether it appears or not. One of the disadvantages of the BoW representation vector is its sparsity, which is caused by the fact that a single document only contains a few vocabulary words. To circumvent the limitations of the BoW methodology, strategies for creating word-embedding Densities Vector Representations (DDRs) using Neural Networks (NNs) have been created. These representations are to some extent capable of encoding specific semantic and syntactic characteristics of words.

The study [18] presented a CNN variant known as BoW-CNN. The same authors developed the Seq-CNN model, which mixes the BoW representations of many words to account for word generation in a sequential order. In the study [19], they proposed utilising a neural network to represent documents while taking the link between phrases into consideration. The neural network initially extracts sentence representations from PDAs using CNN or LSTMs. The neural network then encodes the semantics of the phrases and their underlying connections using a GRULSTM, with the ultimate objective of representing the documents for classification purposes.

An LSTM model was proposed in the work [20] to capture all semantic information in a text. They intend to identify and model generic semantic features while ignoring geographical differences. In their work [21], they propose a hierarchical attention network for predicting the emotional content of publications. When creating a document representation, the model can pay more or less attention to certain words or phrases using two levels of attention mechanisms, one at the word level and one at the sentence level.

The authors of the study [22] constructed an attention-based LSTM network for document-level sentiment prediction. The model for multilingual representation is made up of two LSTMs, each with its own hierarchical structure. This option easily transforms sentiment prediction data from a language with many resources, such as English, to a language with limited resources, such as Chinese, hence improving text classification performance.

For the classification of tweets based on the emotional content of objects whose emotional content depends on the words-expressions mentioned in it, the Adaptive Recursive Neural Network (AdaRNN) model was presented in the study [23] The model feeds the final classification to a softmax classifier using the root node representation of a word as a feature.

The categorization of tweets using characteristics acquired via an unsupervised learning approach was investigated in reference [24]. In this study, it was discovered that employing several vector representations and dictionaries can assist identify texts based on their emotional content.

A proposed attention-based LSTM approach in [25] can identify significant words in a sentence that are related to certain components of the text. Similarly, reference [26] suggests using two LSTM networks that leverage the attention mechanism to improve classifier performance. The researchers enhanced the attention mechanism in the paper [27] so that it now acts in two text boxes to the right and left of the item to be analysed for sentiment.

However, there is recent literature on this subject, including, for example, study [28]. The primary finding of this study is the suggestion of a simple, comprehensive, and successful self-explanatory framework. The additional layer on top of a RoBERTa-large model used for sentiment analysis is the proposed framework's key selling point. This layer generates a neural model of self-explanatory characteristics by collecting information for each text span, assigning a specific weight to each, and then feeding the weighted combination of information to the softmax function for the final prediction.

3.7.2 Summarization

The Transformer serves as the foundation for cutting-edge approaches to summarising in general, and multi-document summarization in particular. The Transformer is an AI architecture that uses the process of attention in order to avoid using recurrency and convolution [28] [29] [30]. In terms of multi-document summarising, [29] use extractive summarization to approximately identify significant information, while the decoder component of the Transformer architecture serves as an abstractive model. This two-stage extractive-abstractive technique is capable of handling extremely long input-output instances and producing meaningful multi-sentence paragraphs.

In [30] the authors propose the use of a neural summarization architecture, the Hierarchical Transformers architecture that can effectively process multiple input documents and produce abstractive summaries. Hierarchical Transformers expand the previously proposed in [31] Transformer architecture with the ability to encode documents in a hierarchical manner by representing cross-document relationships via an attention mechanism. The state-of-the-art PEGASUS model [32], based on pre-training large Transformer-based encoder-decoder models on massive text corpora with a self-supervised objective, is designed for use in both single and multi-document summarization cases. More notably, some other more general-purpose state-of-the-art NLP models, such as BART [33] and T5 [34], can produce comparable results to those of the Pegasus models, mainly in zero-shot and few-shot multi-document summarization task settings. This suggests that unlike single-document summarization, highly abstractive multi-document summarization remains a challenge. A more recent work in multi-document summarization domain is the Multi-Granularity Summarization (MGSum) [35] approach that unifies the extractive and abstractive summarization by utilizing the word representations to generate the abstractive summary and the sentence representations to extract sentences. In this method the PEGASUS model has been used for multi-document summarization. In particular the PEGASUS model has reached SOTA (State-Of-The-Art) on a selected summarization dataset.

The topic of summarization has advanced significantly with the introduction of Large Language Models (LLMs), notably those based on transformer architectures such as GPT-4, BERT, and T5. These models have transformed the capabilities and performance of automatic summarising, pushing the limits of what is possible in both extractive and abstractive summarization tasks. Abstractive summarization, which entails creating new sentences that capture the core of the original text, has improved dramatically using LLMs such as GPT-3 [36] and GPT-4 [37]. These models are pre-trained on massive volumes of data, allowing them to provide logical and context-relevant summaries.

3.8 Lighting and Audio simulations

3.8.1 Lighting simulation

Recent advancements in generative models have introduced diffusion models as a robust approach for image relighting. These models leverage noise-based training processes to learn the transformation of images under varying lighting conditions. Below is a review of the state-of-the-art methods utilizing diffusion models for relighting tasks.

1. Neural Gaffer: Relighting Any Object via Diffusion [1][12]

   Neural Gaffer employs a diffusion model to achieve high-quality 2D relighting of objects. The method involves training on a synthetic dataset called *RelitObjaverse*, which includes 18.4 million rendered images with diverse lighting conditions. The approach integrates Environment Map Rotation and HDR-LDR conditioning to ensure the diffusion model effectively handles the complex relationship between lighting and object appearance. The Environment Map Rotation aligns HDR maps with the camera frame to facilitate learning, while the HDR-LDR conditioning utilizes a dual map approach to maintain lighting detail and balanced exposure. The model architecture integrates an image-to-image latent diffusion model conditioned on input images and lighting maps.

2. Relightify: Relightable 3D Faces from a Single Image via Diffusion Models [1][13]

Relightify uses diffusion models for reconstructing 3D facial BRDF (Bidirectional Reflectance Distribution Function) from a single image, allowing for realistic 3D facial relighting. The model captures diffuse and specular albedo, and normals, generating high-quality 3D avatars that can be rendered under various lighting conditions. The Reflectance Reconstruction jointly recovers facial textures and reflectance for realistic relighting.

3. DiFaReli: Diffusion Face Relighting[33]

DiFaReli addresses face relighting using diffusion models, tackling challenges such as cast shadows, highlights, and unusual makeups without needing 3D or lighting ground truth. The model employs a conditional diffusion implicit model (DDIM) to disentangle and manipulate lighting information effectively. By utilizing its shadow manipulation, the model adjusts the strength and direction of shadows in the relighted images, while its lighting consistency maintains consistent lighting across frames for video relighting.

4. LightIt: Illumination Modeling and Control for Diffusion Models [1][14]

LightIt introduces a method for explicit illumination control in image generation using diffusion models. The model conditions the generation process on shading and normal maps to achieve high-quality, controllable lighting, with a shading estimation module and a control network enabling precise lighting adjustments. The method's key contributions are the explicit control over illumination settings and the high-quality image generation with consistent lighting.

5. IllumiNeRF: 3D Relighting without Inverse Rendering [1][15]

IllumiNeRF offers a novel approach for 3D relighting by leveraging diffusion models instead of traditional inverse rendering. It uses a latent NeRF (Neural Radiance Field) to distill relit images into a 3D representation, allowing for rendering under different lighting conditions. This way the method avoids computationally expensive inverse rendering.

6. IC-Light: Imposing Consistent Light[34]

IC-Light focuses on text-conditioned and background-conditioned relighting of images. This model uses diffusion models to manipulate the illumination of foreground images based on user-specified text prompts or background conditions. The model is able to achieve high performance in various lighting preferences and scenarios.

7. DiffusionLight: Light Probes for Free by Painting a Chrome Ball[35]

DiffusionLight offers a unique technique for estimating lighting in an image by rendering a chrome ball into the scene using diffusion models. This approach leverages the relationship between the appearance of chrome balls and diffusion noise maps to generate high-quality lighting estimates. The method utilizes a fine-tuned diffusion model for HDR light estimation.

8. SwitchLight: Co-design of Physics-driven Architecture and Pre-training Framework for Human Portrait Relighting [1][16]

SwitchLight introduces a human portrait relighting model that combines a physics-guided architecture with a pretraining framework. The utilization of the Cook-Torrance reflectance model enhances the simulation of light interactions with surfaces, while the pretraining masked autoencoder (MAE) framework expands the scale of the training data and enhances the performance in real-world scenarios.

---

[33] https://diffusion-face-relighting.github.io/ (Accessed on 26 July 2024).
[34] https://github.com/lllyasviel/IC-Light (Accessed on 26 July 2024).
[35] https://diffusionlight.github.io/ (Accessed on 26 July 2024).

Funded by the European Union

9. GeoWizard: Unleashing the Diffusion Priors for 3D Geometry Estimation from a Single Image [1][17]
GeoWizard leverages the advantages of diffusion priors to estimate geometric attributes, like depth and normals, improving generalization, detail preservation and efficiency in resource usage. The proposed strategy of breaking down complex data distribution of various scenes to simpler distributions enables the more effective recognition of different scene layouts. This results in high quality depth and normal predictions, which further enhances downstream applications, such as 3D reconstruction and novel viewpoint synthesis.

3.8.2 Audio simulations

3.8.2.1 Current State-of-the-Art:

Existing audio simulation systems and datasets most commonly use:

- **RIR convolution:** This is the calculation of the acoustic response of a room through convolving an anechoic audio signal with the room impulse response function. The room impulse response function is calculated using sinusoid sweep functions or impulse excitation noises (e.g., balloon popping etc.)
- **A-RIR convolution**: This is a similar process that involves the measurement of the room response after the excitation with the sinusoidal function or impulsive noise using a B-FORMAT enabled microphone array (mostly soundfield microphones). This provides four components of the room impulse response in the three axes of the 3D space.
- **Architectural acoustics simulation:** In this family of techniques, the room acoustic response is calculated using computational approaches based on the wave propagation, the geometry and the acoustic attributes of the material of a 3D-modeled room. It is a high-demanding process computationally, that is often approximated using simplified geometrical shapes.
- **Acoustic matching:** It is the most recent advance in room acoustics simulation that is based on deep learning techniques. A model is trained based on anechoic data, and the responses of various spaces having the same data as input. In the end, a model is trained to transform an anechoic audio signal based on a reference audio signal recorded in the target space.

3.8.2.2 SCENE's Technological Innovation:

SCENE enhances current practices through the following technical implementations:

- **Superposition of methods:** The SCENE approach focuses in the extensibility of the data lake by incorporating heterogeneous data. For this reason, a superposition of techniques is deployed that offers a flexibility to contributing users to provide different types of input for the modeling of spaces. Indicators regarding the robustness of every simulation are provided for the end-user.
- **Continuous 3D space live rendering**: The SCENE aims at offering an immersive experience of the acoustic response of a location. For this reason, every location is not characterized by one unique transfer function as it happens in existing dataset. On contrary, an acoustic model is developed to provide a simulation for different positions in the simulated space, based on the existing set of measurements. For every position, binaural rendering takes place for the live playback to the final user.
- **Blind acoustic parameter estimation:** For the AI-fueled acoustic simulation, instead of acoustic matching, SCENE introduces a two-stage approach. On the first stage, several acoustic parameters that characterize the target location are extracted. In the second stage, auralization is performed based on the extracted parameters. This is preferable for two reasons. On the one hand, it provides also measurable metrics of the acoustic attributes of a room from which the user can extract intuitive information without relying on subjective evaluation, and, on the other hand, the method is more robust to audible artifacts that can be produced during end-to-end acoustic matching

Funded by
the European Union

3.8.2.3 SCENE's Technological Stack:

The main technologies which participate in the development and deployment of SCENE Audio and Lighting simulation tools are:

- **Ambisonics:** Ambisonics are the industry standard for immersive 3D audio experiences. They provide the B-FORMAT encoding that captures all the needed information for the reproduction based on different configuration (azimuth, elevation angle, position, etc.). Ambisonics are used in the SCENE approach both in the capturing (using a portable soundfield microphone array) and in reproduction.
- **Resonance Audio:** Spatial audio rendering and playback is very important to address a target audio. Several options were considered for this, but the JavaScript sdk of Resonance Audio was preferred, offering web-based reproduction on the browser, that is considered inclusive to all users.
- **PyTorch:** PyTorch is a deep learning library for python that facilitates the process of training, evaluating and deploying models. In the case of the audio simulation module, it is used to train models for data-driven blind acoustic parameter estimation.

3.9 UWB-based Tracking System

The Ultra-WideBand (UWB) is a wireless technology that enables precise localization with less than 20 centimetres accuracy by measuring the Time-of-Flight (ToF) of radio signals between UWB devices.
The UWB was defined and standardized by IEEE 802.15.4a in 2007 and later by IEEE 802.15.4z in 2020. In particular, the last standard increased the integrity and accuracy of distance measurements. The tracking solution involves UWB devices (anchors and tags) deployed in an area of interest. Mobile tags communicate with fixed anchors to determine their position. Specifically, tags and anchors measure round-trip ToF between themselves by using a method called Two-Way Ranging (TWR). After that, a trilateration algorithm estimates tags' position. In the SCENE's context, mobile tags could be wearable devices for tracking the actors.
Other ranging methods can be applied such as TDoA (Time Difference of Arrival), where tags broadcast "blink" messages, and anchors timestamp them. PDoA (Phase Difference of Arrival) is another ranging method similar to TDoA but it uses phase differences for positioning.
Some important features of the UWB technology are summarized in Figure 8 (from Qorvo).

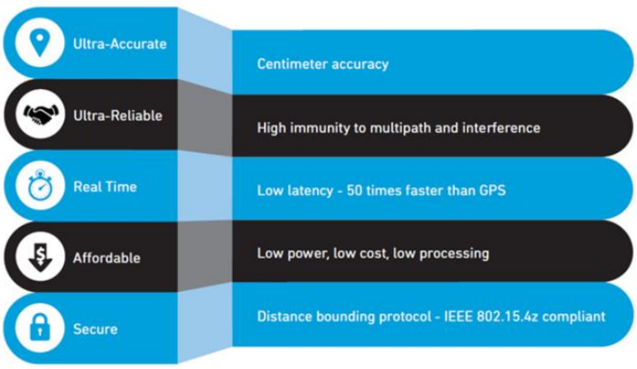


Figure 8. Main UWB features – source from Qorvo.

The UWB has been adopted in various applications such as logistics, healthcare, manufacturing, and transportation for real-time positioning and tracking. Indeed, UWB is highly accurate for both indoor and outdoor environments, making it valuable for asset tracking and people localization.

3.9.1 Current Start of the Art

The UWB has become a mature technology since the new IEEE 802.15.4z standard. Moreover, there are several consolidated platforms commercially available on the market. Some of them are briefly presented as follows.

- **Qorvo** (ex Decawave) UWB technology has been applied within mobile, automotive, industrial and consumer IoT markets and applications. The UWB technology pioneered by Decawave (acquired by Qorvo in 2020) enables centimeter accuracy distance / location measurement and secure low-power, low-latency data communication. In particular, Qorvo delivers a comprehensive toolkit to ease the development of micro-location solutions. They also design and provide UWB-based chips (i.e., integrated circuits - ICs), which implement the full UWB protocol stack, capable of operating with coin cell batteries for various types of applications.
- **Sewio** provides a Real-Time Location System (RTLS) for precise indoor tracking mainly used for industrial applications. In particular, the Sewio platform provides both hardware and software solutions allowing to gain accurate and actionable data leading to a productive, cost-effective and safe industries.
- **Localino** offers localization systems based on the UWB technology. These are characterized by a simple modular design and low infrastructure requirements. These have been employed for industrial, retail, logistics and sports applications, among others.
- **Pozyx** platform brings indoor and outdoor positioning data together to provide full asset visibility and location-based insights mainly for logistics and manufacturing applications. It facilitates warehouse and inventory control, keeps track of returnable packaging and orders, and slashes lost asset costs.
- **Eliko** platform allows to precisely track vehicles, goods and people in real-time with a reliable indoor positioning system. It has been designed for digital solutions that depend on location data. It is worth mentioning the Entertainment Case Study that is focused on accurate tracking of actors on stage with the objective of automatically activating microphones, sound, and lighting effects at the precise moment required. By exploiting this solution, the actors focus only on acting and not on the technology.

It is worth remarking that all the above mentioned UWB platforms are based on the UWB chips designed and produced by Qorvo.

3.9.2 SCENE's Technological Innovation

LINKS has developed its own UWB platform called **Artemis** mainly used in research projects, within which it is further developed and enhanced, and to strength its technology transfer service.
Artemis is a large-scale localization solution designed to track a wide number of tags in areas of any size, aiming to support Industry 4.0 applications as well as Robotics applications in indoor environments. The large-scale feature is enabled by a protocol, named Time-Division Multiple Access (TDMA), implemented in all UWB devices (i.e., tag, anchor, GW), while ranging measurements are based on the TWR method (mentioned in section 3.9.1).
We have compared Artemis with the above presented UWB platforms. Artemis provides similar performance to the Qorvo solution, namely MDEK 1001. Specifically, the location accuracy is about 20 cm, the deployment of the UWB infrastructure is relatively simple thanks to the TWR method, the localization capacity is 150 positions per second (each tag performs TWR with four anchors) and the maximum tracking frequency is 10 Hz.

Table 2: Comparison of the technologies available for tracking systems

Funded by
the European Union

FEATURES	Artemis LINKS	Qorvo (Decawave MDEK1001)	Sewio	Localino	Pozyx	Eliko
Localization accuracy < 20 centimetri	✓	✓	✗	✗	✗	✗
Simple ranging method (e.g., TDoA)	✗	✗	✓	✓	✓	✓
Simple deployment (e.g., no Internet connection is required for all anchors)	✓	✓	✗	✗	✗	✗
Localization capacity	150 loc./sec	150 loc./sec	N.A.	N.A.	1000 loc./sec	N.A.
Localization frequency (per tag)	10 Hz	10 Hz	20 Hz	100 Hz	75 Hz	40 Hz

Similar to Eliko use case, the Artemis UWB-based tracking solution will be integrated in the SCENE platform to track actors' position to automatically activate some effects, such as lighting and sound effects, during the production phase.

The performance of Artemis could be enhanced by combining both UWB ranging and IMU (Inertial Measurement Unit) data overall providing a robust and more accurate location estimation in case the indoor environment is affected by interference from large obstacles that cause intermittent UWB connectivity. Thus, by exploiting data from accelerometer and gyroscope sensors (integrated in the IMU), it would be possible to track the actors even though the UWB is not available for short interval of time.

3.10 Distribution Engine

This section describes the current market solutions distribution engines, analysing their ad-vantages and disadvantages and comparing them with the Distribution Engine and Media Asset Manager tools offered by SCENE.

3.10.1 RightsTrade

RightsTrade[36] is an online marketplace that streamlines the licensing and distribution process for film, television, and digital media. It aims to connect content owners with buyers worldwide, offering a platform for negotiating and finalizing deals without intermediaries.

RightsTrade offers a variety of features, including a comprehensive catalogue of media content, tools for direct negotiations between buyers and sellers, and automated licensing processes. The platform provides secure digital contracts and payment processing, along with detailed analytics to track licensing activity.

Despite its advantages, RightsTrade has some limitations. It does not utilize blockchain technology, which can enhance security and transparency. The platform may also require users to have a certain level of expertise in media licensing, potentially making it less accessible for newcomers. Additionally, while RightsTrade streamlines many processes, it may not eliminate the need for traditional legal oversight in complex deals.

3.10.2 MILC

MILC (Media Industry Licensing Content)[37] is a blockchain-based platform that aims to revolutionize the media licensing and distribution industry by leveraging the security and transparency of blockchain technology. The platform uses Hyperledger Fabric to securely register contract data and facilitate smart contracts, ensuring compliance with international standards. MILC also introduces its own cryptocurrency (MILC) to streamline transactions, making the licensing process more efficient and secure.

MILC offers a range of features designed to enhance media licensing and distribution. It provides secure and immutable records of contracts through blockchain, ensuring that all transactions are transparent and tamper-proof. The platform utilizes smart contracts to automate licensing agreements, reducing the need for

[36] https://rightstrade.com/ (Accessed on 26 July 2024).
[37] https://milc.global/ (Accessed on 26 July 2024).

Funded by the European Union

intermediaries, and minimizing administrative overhead. Additionally, MILC supports cryptocurrency transactions, offering a fast and cost-effective payment solution. Its growing catalog of media content is accessible to a global audience, with tools for detailed analytics and reporting.
While MILC brings significant advantages through blockchain technology, it also has some limitations. The use of cryptocurrency for transactions can be a barrier for users unfamiliar with digital currencies. The platform is also more tailored to VR, with a big focus on gaming.

3.10.3 LBRY

LBRY[38] is a decentralized platform that aims to disrupt the traditional media distribution model by leveraging blockchain technology. It provides a decentralized marketplace for sharing and publishing content, where creators can upload their work and retain complete control over distribution and monetization. LBRY uses a blockchain-based protocol to ensure transparency and security, making it an attractive option for content creators seeking independence from traditional media gatekeepers.
LBRY offers a robust set of features tailored to the needs of content creators and consumers. The platform's decentralized nature ensures that content is distributed across a peer-to-peer network, reducing the risk of censorship, and enhancing content security. LBRY uses a unique token (LBC) to facilitate transactions, allowing creators to monetize their work directly from their audience. The platform also includes detailed analytics and reporting tools to help creators understand their audience and optimize their content strategy. Additionally, LBRY provides an open-source protocol, enabling developers to build applications on top of the LBRY network.
Despite its innovative approach, LBRY faces several limitations. The reliance on cryptocurrency (LBC) can be a barrier for users unfamiliar with digital currencies, potentially limiting the platform's adoption. The platform's user interface and overall user experience may not be as polished as some traditional, centralized alternatives and it’s also not customizable to the user needs.

3.10.4 Brightcove

Brightcove[39] is a leading cloud-based video platform designed to facilitate the hosting, management, and distribution of video content. It is widely used by enterprises, media companies, and marketers to deliver high-quality video experiences to audiences across various devices and platforms. Brightcove’s platform offers robust video solutions, including live streaming, video-on-demand (VOD), and monetization tools, making it a comprehensive solution for organizations looking to leverage video content effectively.
Brightcove provides a wide array of features that cater to the diverse needs of video content creators and distributors. The platform supports advanced video processing capabilities, including encoding, transcoding, and adaptive bitrate streaming to ensure optimal video quality and performance. Brightcove’s video cloud offers extensive analytics and reporting tools, allowing users to track viewer engagement and measure the effectiveness of their video content. The platform integrates seamlessly with various third-party applications and services, providing flexibility and scalability. Additionally, Brightcove supports monetization through advertising, subscription models, and pay-per-view options, enabling content creators to generate revenue from their video content.
Despite its comprehensive feature set, Brightcove has some limitations. The platform can be cost-prohibitive for smaller businesses and individual content creators due to its pricing structure, which is geared towards enterprises and large media companies. Brightcove’s extensive features and capabilities can also introduce complexity, potentially requiring a steeper learning curve and more technical expertise to fully utilize the

[38] https://lbry.com/ (Accessed on 26 July 2024).
[39] https://www.brightcove.com/en/ (Accessed on 26 July 2024).

Funded by the European Union

platform. Furthermore, while Brightcove offers robust video solutions, it may not be as flexible or customizable as some open-source alternatives, limiting the ability to tailor the platform to specific needs. Finally, users may face challenges with integration and customization without sufficient technical support, which can be a drawback for those seeking highly specialized video solutions.

3.10.5 Market analysis and comparison with SCENE

Table 3 provides a comprehensive comparison of various tools used in media licensing and distribution, including RightsTrade, MILC, LBRY, and Brightcove, together with the SCENE platform to be developed during the project. Each tool is evaluated across several key functionalities: user interface, scalability, license management, media distribution, decentralized license agreements, ease of implementation, and customizable distribution platform. This comparison aims to highlight the strengths and limitations of each tool, offering insights into their suitability for different business needs and technical environments. The table illustrates how the SCENE platform, powered by both the Distribution Engine and Media Asset Manager, can uniquely position itself in the market by offering advanced features that address the limitations of current solutions. Most media distribution solutions do not handle media asset management and/or license management, requiring additional services to be used.

Table 3: Comparison between existing Distribution Engines and the SCENE platform

Platform	User Interface	Scalability	License Management	Media Distribution	Decentralized License Agreements	Ease of implementation	Customizable Distribution platform
RightsTrade	Intuitive and user-friendly	Suitable for all sizes	Comprehensive	No	No	High	No
MILC	User-friendly but requires crypto knowledge	Suitable for all sizes	Comprehensive	Advanced	Yes	Medium	Only on request
LBRY	User-friendly but technical	Suitable for all sizes	Comprehensive	Advanced	Yes	Medium	No
Brightcove	Intuitive and user-friendly	Suitable only for large enterprises	Comprehensive	Advanced	No	High	Yes
SCENE	Intuitive and user-friendly	Suitable for all sizes	Comprehensive	Advanced	Yes	High	Yes

3.11 Recommendation Systems

Recommender systems are sophisticated algorithms designed to help users discover items of interest by filtering and suggesting information tailored to their preferences. These systems analyse user data and leverage various techniques, such as collaborative filtering, content-based filtering, and hybrid approaches, [2] [1]

Funded by
the European Union

to predict and recommend products, services, or content that align with individual tastes and needs. Commonly used in e-commerce, streaming services, and social media platforms, recommender systems enhance user experience by providing personalized recommendations, thereby increasing user engagement and satisfaction.

Collaborative filtering is a cornerstone technique in recommender systems, leveraging the wisdom of the crowd to provide personalized suggestions. The state-of-the-art in collaborative filtering continues to evolve, integrating sophisticated machine learning methods to enhance performance and scalability. Two notable advancements in this field are "Latent Relational Metric Learning via Memory-based Attention for Collaborative Ranking" (LRML) [1][2] and "The Minority Matters: A Diversity-Promoting Collaborative Metric Learning Algorithm" (DPCML) [1][3].

LRML is a neural architecture designed for collaborative ranking with implicit feedback, presenting a novel approach to metric learning for recommendations. Traditional metric learning approaches often rely on simple push-pull mechanisms between user and item pairs, which can be geometrically inflexible. LRML addresses this limitation by learning latent relations that describe each user-item interaction. This method enhances modelling capabilities and scalability for a large number of interactions.

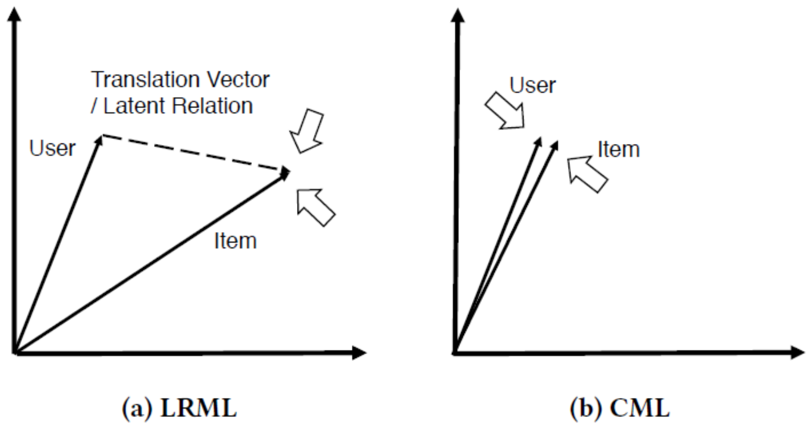


Figure 9: Geometric Comparison of LRML and CML for modelling User-Item Relationships in metric space. LRML uses a Translation Vector /Latent Relation. (Yi, Luu, & Siu, 2018)

DPCML addresses a key challenge in collaborative filtering related to user representation. In typical methods, a unique user representation can induce preference bias, particularly when users have multiple categories of interests and item category distributions are imbalanced. DPCML proposes a solution by incorporating multiple sets of representations for each user.

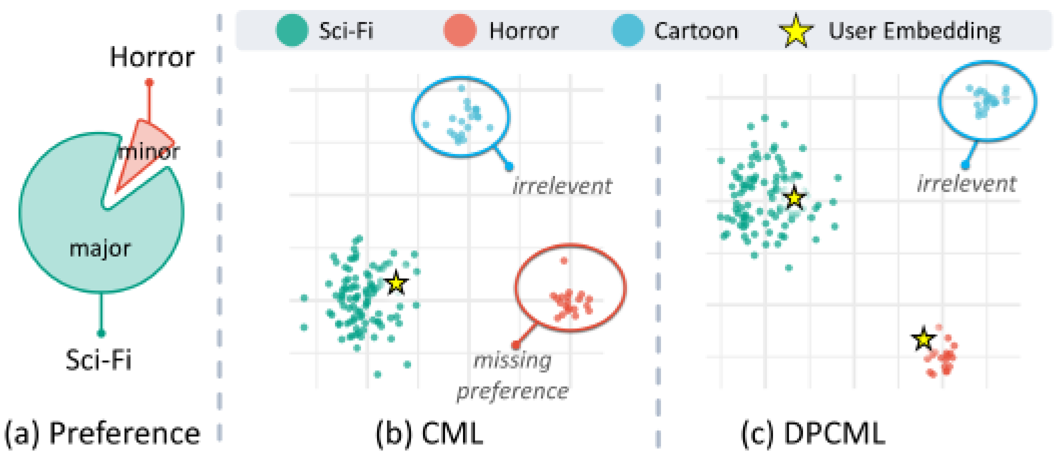


Figure 10: An illustration that shows the benefits of DPCML when a user has multiple categories of preferences. If two categories are separated, traditional methods have lesser accuracy. (Shilong, et al., 2022)

Similarly, the paper "Effective metric learning with co-occurrence embedding for collaborative recommendations" [3] describes a model that considers the global statistical information of item to item and user to user pairs. It involves a co-occurrence embedding that regularizes the metric learning model.

The hybrid approach in recommender systems combines multiple recommendation techniques, such as collaborative filtering, content-based filtering, and sometimes even knowledge-based methods, to leverage the strengths of each and mitigate their individual limitations. This integration enhances recommendation accuracy and robustness by utilizing diverse data sources and algorithms, leading to a more comprehensive understanding of user preferences. Hybrid systems can effectively address issues like the cold-start problem, where limited user data hampers accurate recommendations, by using content-based methods to supplement sparse collaborative data.

A state-of-the-art hybrid model is being described in the following paper: "Multi-Modality is All You Need for Transferable Recommender Systems" [4]. It is a fusion of a simple collaborative filtering methodology and a multimodal pipeline. Using text encoders (BeRT) and image encoders (CLIP), it creates a latent representation of every item. This methodology enhances the resulting model, increasing the accuracy. Additionally, it mitigates the problem of non-accurate results when a user has limited interactions (cold start problem). In the same vein, the paper "Where to Go Next for Recommender Systems? ID- vs. Modality-based Recommender Models Revisited" [5] introduces a multi-modal recommender system that uses images or text data, but not simultaneously.

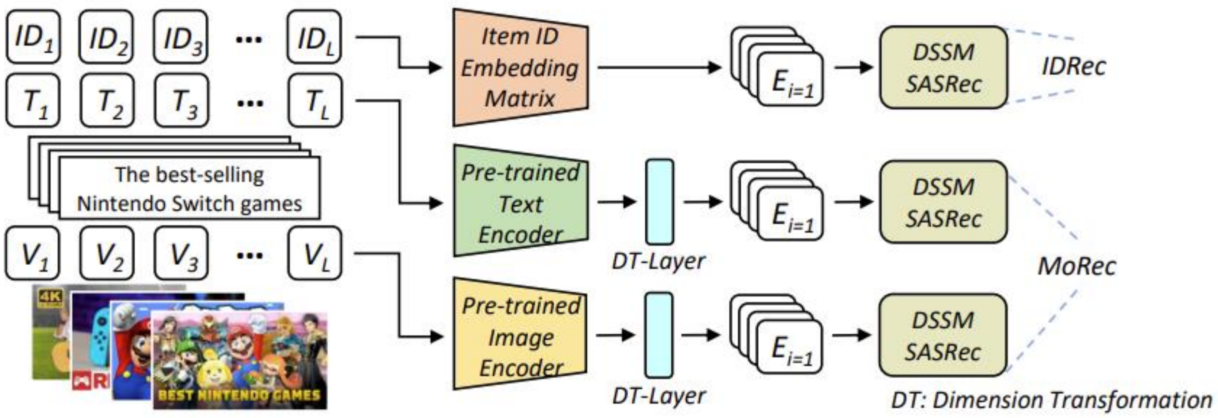


Figure 11: Comparison between an item ID based recommender system and a vision or text modality recommender system. Using pretrained Text and Image encoders, the modal recommender system leverages more information about the items. (Zheng, et al., 2023)

The proposed model by the SCENE project is based on the state-of-the-art "Latent Relational Metric Learning" method. This methodology is being extended to use metadata about the movies in an effort to increase the accuracy and effectively mitigate the cold start problem. Metadata like this include information about the genre of a movie, the available budget and more.

3.12Quality metrics & post-production effects

Video quality metrics have become widely available in recent years. The most comprehensive suite of measurement tools is currently probably the MSU Video Quality Measurement Tool, containing both reference and non-reference metrics. Among the metrics supported are SSIM, MS-SSIM, PSNR, VMAF, NIQE, MSAD, MSE, VQM, Blurring and Blocking. The most popular metrics, VMAF, PSNR and SSIM are also available in mainstream tools like FFMPEG.

These tools have been primarily used in recent years for the development and evaluation of video co-decs, such as HEVC (H.265) or VVC (H.266). To the best of our knowledge, they have not yet been used for the evaluation of post-production effects or 3D model quality.

While 3D models are common in the production of videos, their primary purpose is to create elements not previously present in a scene. Hence reference metrics tools are not suitable for most evaluations of model quality. Model quality is commonly measured based on technical parameters, such as mesh density, mesh quality, texture fitting and quality, as well as performance analysis.

Visual quality of the resulting rendering of the 3D models is commonly determined by looking at the resulting footage.

As the primary purpose of 3D modelling and rendering in SCENE is to provide an accurate representation of the potential shooting location, the project is in the atypical position that models should be able to match photos and videos taken on location as closely as possible. This allows innovative uses of reference metrics, originally devised for codec development, in innovative ways, which go beyond the state of the art.

4 Software and Hardware Specification of SCENE's Modules

An Excel template (Figure 12) was sent to the technical partners for the efficient collection of software (SW) and hardware (HW) requirements. This template defines ten fields for SW needs and seven for HW requirements, guaranteeing an organized and uniform approach across all submissions. To make things easier and clearer, full examples and descriptions were supplied for each area. This allowed all stakeholders to completely understand the expectations and needs.

Each technical partner was responsible for describing the areas pertinent to SCENE's components and needs, ensuring that all parts of the system were adequately covered, while defining the technical requirements of the SCENE platform that will integrate all components implemented within WP3 and WP4. This approach enabled the detailed collection of needs for each component, including those pertaining to Generative AI. By gathering this information in this manner, a solid basis for the system's growth and integration has been set.

6	Software Requirements			
7	Field	Details	Example	Description
8	Operating System		Windows / Linux	Specify the operating systems the software is compatible with.
9	Database Details		Mysql 5.6, MongoDB 4.2	Mention the type and version of the database used.
10	Exposed APIs		Restful web-services	List any APIs the software exposes for integration.
11	Exchanged Data Format		Python, Java, C++	Define the format in which data is exchanged between systems.
12	Source Code Programing Language		JSON	Indicate the primary programming language used in the software's source code.
13	AI Frameworks		Tensorflow, Pytorch	Specify any AI or machine learning frameworks utilized.
14	User Interface Framework (if applicable)		React.js, Vue.js	Mention the framework or library used for the user interface.
15	Security Protocols		OAuth 2.0, SSL/TLS	List security protocols or standards the software adheres to.
16	Integration Tools		Zapier, AWS Lambda	Specify tools or platforms the software can integrate with.
17	Version Control System		Git (GitHub), Mercuria	Indicate the system used for tracking changes in the software's source code.
18				
19	Hardware Requirements			
20	Field	Details	Example	Description
21	Device Type		Laptop, Server, IoT Device	Specify the category or type of the hardware device.
22	Processor		Intel i7-9750H, ARM Cortex-A53	Indicate the type and speed of the processor.
23	Memory		16GB DDR4	Mention the amount and type of RAM.
24	Storage		512GB SSD, 2TB HDD	Specify the type and capacity of storage.
25	Connectivity		Wi-Fi 6, Bluetooth 5.0, Gigabit Ethernet	List the connectivity options the hardware supports.
26	Power Supply		65W AC Adapter, 5000mAh Battery	Indicate the power source and its capacity.
27	Graphics Capabilities		NVIDIA GeForce GTX 3090Ti	Specify the graphics processing capabilities.

Figure 12: Template distributed for the software and hardware collection

The following sections present the SW and HW requirements collected for each component.

4.1 Data Lake

4.1.1 Software Requirements and Hardware Justification

Specific software and hardware configurations are essential to operate our Data Lake tools effectively. For instance, the **Data Lake (MinIO)** requires a robust server environment due to its role as a high-performance, S3-compatible object storage system. MinIO is designed to handle large-scale data storage and management, necessitating significant memory and processing power. Therefore, a server with at least 32GB of RAM and multi-core processors (e.g., 8-core CPU) is recommended. This configuration ensures efficient handling of metadata operations and concurrent data access by multiple users. Additionally, ample storage space, preferably using SSDs for faster data retrieval, is crucial to support the high I/O operations that MinIO demands. For the SCENE project, we will not require that fast access and have used SATA disks with automatic (daily) backup.

It is not mandatory for the Data Lake, but substantial resources will be required if it is necessary to incorporate a Data Warehouse (**Apache Hive**) in a second phase. A typical deployment of Apache Hive would benefit from a server with 64GB of RAM to optimise the in-memory computations performed by the Hive execution engine. Additionally, the server should have a high-throughput network interface to manage large data transfers efficiently. These hardware specifications are justified by the need to perform complex SQL-like queries on petabyte-scale datasets, ensuring low latency and high throughput. For the SCENE project, we will try to accommodate resources that meet real needs; thus, the initial resources will probably be lowered.

The **Operational Tools (Scheduler)** is another essential component that requires specific hardware to function optimally. This tool is responsible for scheduling and orchestrating various tasks within the data lake ecosystem, ensuring that data ingestion, processing, and querying tasks are executed promptly and efficiently. Given its role in managing numerous concurrent tasks and workflows, the scheduler requires a server with at least 16GB of RAM and a quad-core processor. This ensures the scheduler can handle multiple threads and processes without degradation. Additionally, reliable storage and backup solutions are necessary to maintain the scheduler's logs and configurations, preventing data loss and ensuring continuity in operations. The usage of this tool depends on the number of models or predictive algorithms that are included in the system to perform data analysis; currently, SCENE tools are independent (stand-alone) applications, but some parts might be modularised, as well as future components after the project.

4.1.2 Limitations and Constraints

Several limitations and constraints have been identified during the development and deployment of our Data Lake tools. One significant constraint is the handling of personal and sensitive data. For instance, while ingesting data into the **Data Lake**, ensuring compliance with data privacy regulations such as GDPR is crucial. This necessitates implementing data anonymization and encryption mechanisms, which can introduce additional overhead and complexity into the system. These security measures can impact the performance and resource requirements, as encrypting and decrypting data is computationally intensive and demands more processing power and memory. The SCENE tools can also handle the anonymization process before using the Data Lake to offload the Data Lake requirements; this has not yet been decided and will be subject to analysis in the upcoming months. There will probably be an intermediate approach as the ingestion phase of the data lake typically combines features of the data lake with features of (some of) the SCENE tools.

Another limitation is related to integrating different data lake components, such as **MinIO, Apache Hive,** and **Trino**. Each element has its own dependencies and configuration requirements, which can lead to compatibility issues. For example, ensuring that the data formats and schemas are consistent across these tools requires meticulous configuration and testing. This complexity can impact the ease of deployment and maintenance, necessitating a well-coordinated approach to manage version compatibility and interoperability among the various data lake modules. We are incrementally conducting a docker-compose approach to avoid complex Kubernetes deployments.

4.2 Knowledge Graph Management & Ontology Formulation

The following table presents the technical specifications, in terms of software and hardware requirements, for the Knowledge Graph Management & Ontology Formulation tool that will be integrated within the SCENE platform.

Table 4: Software and hardware requirements of the Knowledge Graph Management & Ontology Formulation

Module	Hardware Requirements	Software Requirements	Security and Privacy	Dependencies

Ontology formulation & knowledge graph manager tool	- Server - Intel i7-9750H - 16GB DDR4 - 512GB SSD, 2TB HDD	- Linux - Python - Git	- GDPR compliance	- Python 3.8 or later - Conda - PyTorch 1.8.0 or later

4.3 Media Asset Manager and Distribution Engine

The table below shows the required software and hardware for the Media Asset Manager and the Distribution Engine modules.

Table 5: Software and hardware requirements of the Media Asset Manager and Distribution Engine

Module	Field	Software Requirements	Field	Hardware Requirements
Media Asset Manager and Distribution Engine	Operating System	Linux	Device Type	Azure Kubernetes Service v1.26.3
	Database Details	MongoDB 4	Processor	3x 4 vCPU
	Exposed APIs	Restful API	Memory	3x 16 GiB
	Exchanged Data Format	JSON	Storage	20GB + 1TB External S3 storage
	Source Code Programing Language	Javascript / Typescript	Connectivity	Ethernet
	AI Frameworks		Power Supply	
	User Interface Framework (if applicable)	React.js	Graphics Capabilities	
	Security Protocols	SSL/TLS	Device Type	Azure Kubernetes Service v1.26.3
	Integration Tools		Processor	3x 4 vCPU
	Version Control System	Gitlab	Memory	3x 16 GiB

To effectively run the Media Asset Manager and Distribution Engine tool, the recommended hardware configuration includes a server environment with Azure Kubernetes Service, which provides scalable and reliable cloud-based infrastructure. The processing needs are met with a setup of 3x 4 vCPUs, ensuring sufficient computational power for media processing tasks. With 3x 16 GiB of RAM, the system supports concurrent operations and large-scale data handling, also required when handling high resolution media. Storage requirements include 20GB of internal storage along with 1TB of external S3 storage to accommodate extensive media assets and metadata.

4.4 3D Reconstruction

The table below shows the required software and hardware for the 3D model reconstruction module.

Table 6: Software and hardware requirements of the 3D Reconstruction

	Field		Field	

Funded by the European Union

Module		Software Requirements		Hardware Requirements
3D Reconstruction	Operating System	Windows / Linux	Device Type	Laptop / Desktop / Server
	Database Details	Graph database	Processor	Intel i7 11th Gen.
	Exposed APIs	Restful web-services	Memory	16GB DDR4
	Exchanged Data Format	.fbx, .obj, Json, Json-LD	Storage	256 GB
	Source Code Programing Language	Python, C++, C#	Connectivity	Wi-Fi 6 and / or Ethernet
	AI Frameworks	Tensorflow	Power Supply	85W - 150W
	User Interface Framework (if applicable)	Unity3D (Possibly to have in Unreal also.)	Graphics Capabilities	NVIDIA GeForce RTX 3080Ti
	Security Protocols	OAuth 2.0, SSL/TLS		
	Integration Tools	N/A		
	Version Control System	GitHub		

For performing the 3D model reconstruction, a laptop/desktop equipped with minimum 16 GB RAM with sufficient disk space (e.g., 256 GB), a moderate processor (e.g., Intel i7 11th Gen.) with NVIDIA Graphics card (e.g., GeForce 3080 Ti) for compiling CUDA-based dependencies in 3D Gaussian Splatting or NeRF implementations and a stable internet connectivity is required. These requirements are based on the rather demanding processing pipeline for the automatic 3D model rendering/reconstruction tools e.g., Gaussian splatting, NeRF. Moreover, other 3D modelling tools such as Blender also require such hardware requirements as the base requirement. Similarly, for developing and deploying the 3D models, we need laptops/desktops with Windows/Linux operating systems. The development language includes Python, C++ and also C# for implementing the tools and technologies such as NeRF, 3D Gaussian splatting, Photogrammetry. The reconstructed 3D models can be exchanged in suitable format, e.g., .fbx; which can be later viewed in UI developed, for example, using Unity.
At this stage of development, we did not yet encounter any limitations or constrains. We will report on any limitations or constraints which might occur during the continuation of the tools' development in the upcoming deliverables.

4.5 Blockchain Technologies

The following table presents the technical specifications for the blockchain framework module within the SCENE project.

Table 7: Software and hardware requirements of the Blockchain Technologies

Module	Hardware Requirements	Software Requirements	Security and Privacy	Dependencies
Blockchain Technologies	**Minimum**: - 4 vCPU - 8GB RAM - 50GB Storage	- **Hyperledger Fabric v2.3**: The base plat-	- **Data Privacy**: Leveraging Hyperledger Fabric's	- **Hyperledger Fabric SDKs** (for Golang language): Required for

Funded by the European Union

	Recommended: - 8 vCPU or more - 16GB RAM or more - 100GB Storage or more (Scalable based on network size and transaction volume)	form for the permissioned blockchain network, providing core features like consensus mechanisms, chaincode execution, and membership services. - **Golang**: The programming language used to develop custom chaincodes for NFTs and Ricardian Contracts, as well as the HLF Gateway API service. - **Docker**: Containerization platform for packaging and deploying components, ensuring portability and ease of management. - **IPFS Client/Node**: Software for interacting with the IPFS network and storing/retrieving media assets.	channel architecture to isolate data between different organizations and ensure confidentiality. - **Access Control**: Implementing strict access control mechanisms within the chaincode and HLF Gateway to restrict sensitive operations to authorized parties. - **Encryption**: Employing encryption techniques (e.g., AES) to protect data both at rest and in transit, particularly for sensitive information like private keys and personally identifiable information (PII). - **Auditability**: Maintaining a transparent and immutable record of all transactions on the blockchain for auditing and compliance purposes. - **Regular Security Audits**: Conducting regular security audits and vulnerability assessments to identify and address potential risks.	application development and interaction with the blockchain network.

The table details the specifications for the blockchain framework module within the SCENE project. The hardware requirements outline the minimum and recommended resources for running the blockchain network, which are scalable based on the network's size and transaction load. The software requirements specify the core technologies utilized, including Hyperledger Fabric as the blockchain platform, Golang for chaincode (Smart Contract) development, Docker for containerization, and IPFS for decentralized storage. Security and privacy measures are paramount in the Blockchain framework. The project leverages Hyperledger Fabric's channel architecture for data isolation, implements strict access controls, and employs encryption for data protection. Auditability is ensured through the transparent and immutable nature of the blockchain, and regular security audits are conducted to mitigate risks. Finally, the dependencies section lists the necessary software components for application development and interaction with the blockchain network, including Hyperledger Fabric SDKs. Optional components like IPFS Cluster and external databases are also mentioned to enhance system availability and manage large datasets.

4.6 Location Scouting

The following table presents the technical specifications, in terms of Software and hardware requirements, for the Location Scouting tool that will be integrated within the SCENE platform. Due to the integration of the generative AI Large Language models (LLMs) in this tool, and connect it also with the Audience building tool and the Recommendation system, the hardware requirements of the server required are intense.

Table 8: Software and hardware requirements of the Location scouting tool

Module	Hardware Requirements	Software Requirements	Security and Privacy	Dependencies
Location scouting tool	- Server - Intel i7-9750H - 32GB DDR4 - 512GB SSD, 2TB HDD - Wi-Fi 6, Bluetooth 5.0, Gigabit Ethernet - 65W AC Adapter - NVIDIA GeForce GTX 3090Ti	- Linux - MySQL database - Milvus database - Restful webservices - JSON - Python - to be defined - Angular - OAuth 2.0, SSL/TLS	- RESTful API with HTTPS - SSL/TLS encryption for server communication - GDPR compliance	- Python 3.8 or later - Conda - PyTorch 1.8.0 or later

4.7 Audience Building (AB)

The following table presents the technical specifications, in terms of Software and hardware requirements, for the Audience Building tool that will be integrated within the SCENE platform.

Table 9: Software and hardware requirements of the Audience Building tool

Module	Hardware Requirements	Software Requirements	Security and Privacy	Dependencies

Audience Building (AB) tool	- Server - Intel i7-9750H - 32GB DDR4 - 512GB SSD, 2TB HDD - Wi-Fi 6, Bluetooth 5.0, Gigabit Ethernet	- Linux - MySQL - Restful web-services - JSON - Python - not defined yet - Angular - OAuth 2.0, SSL/TLS - not defined yet - Git	- RESTful API with HTTPS - SSL/TLS encryption for server communication - GDPR compliance	- Python 3.8 or later - Conda - PyTorch 1.8.0 or later

4.8 Lighting and Audio Simulation

4.8.1 Lighting simulation

The following table presents the technical specifications, in terms of Software and hardware requirements, for the Lighting simulation that will be integrated within the SCENE platform.

Table 10: Software and hardware requirements of the Lighting simulation tool

Module	Hardware Requirements	Software Requirements	Security and Privacy	Dependencies
Lighting Simulation	- High-end GPU with a minimum of 12GB VRAM - Multi-core CPU processor - At least 32GB RAM - SSD storage for faster data I/O operations - Effective cooling system and stable power supply	- Operating system: Linux (Ubuntu 20.04 or later) or Windows (10 or later) - CUDA 11.8 or later GPU acceleration framework with compatible cuDNN version	- RESTful API with HTTPS - SSL/TLS encryption for server communication - GDPR compliance	- Python 3.8 or later - Conda - PyTorch 1.8.0 or later

4.8.2 Audio simulations

For the audio simulation tool, PyTorch is needed for model training and deployment, and the Resonance Audio JavaScript SDK is the core of the web-based functionality. It is important to have a processor for more than 8 cores and 16Gb RAM for model execution, while 2TB of HDD are considered sufficient for the storage of models, source audio files and sample audio files that are needed for the simulation.
A slight constraint on the audio simulation tool concerns the hardware requirements of the user, where headphones are preferred for a better playback experience of the binaural rendering. This is indicated in the GUI, so that the users are aware. The tool is functional without headphones as well.

4.9 UWB-based Tracking System

The following table presents the technical specifications, in terms of HW and SW requirements, for the UWB Tracking Solution module that will be adopted within the SCENE project. Specifically, the information listed in the table pertains only the "Tracking Manager" component (refer to D2.6 - Reference Architecture.R1). The specifications of the UWB-based Devices (i.e., Tags and Anchors) have not been included as these are custom devices provided LINKS.

Table 11: Software and hardware requirements of the UWB-based Tracking System

Module	Field	Software Requirements	Field	Hardware Requirements
UWB Tracking Solution	Operating System	Linux, Ubuntu server 22.04	Device Type	Laptop or mini-PC (e.g., Raspberry PI 4)
	Database Details	N/A	Processor	At least a 4-core CPU
	Exposed APIs	MQTT	Memory	16GB DDR4
	Exchanged Data Format	JSON	Storage	at least 20 GB SSD
	Source Code Programing Language	C++	Connectivity	IP connectivity via Wi-Fi or Gigabit Ethernet to the UWB Gateways.
	AI Frameworks	N/A	Power Supply	
	User Interface Framework (if applicable)	Qt Creator	Graphics Capabilities	N/A
	Security Protocols	clients can subscribe to the localization topic by accessing the MQTT broker with username and password. No specific encryption is used at moment.		
	Integration Tools	N/A		
	Version Control System	GitLab. It is used only internally in LINKS. Not externally exposed		

4.10 Recommendation System

The following table presents the technical specifications, in terms of Software and hardware requirements, for the recommendation system that will be integrated within the SCENE platform.

Table 12: Software and hardware requirements of the recommendation system

Module	Hardware Requirements	Software Requirements	Security and Privacy	Dependencies

Funded by the European Union

Recommendation system	- Server - Intel i7-9750H - 32GB DDR4 - 512GB SSD, 2TB HDD - Wi-Fi 6, Bluetooth 5.0, Gigabit Ethernet - NVIDIA GeForce GTX 3090Ti	- Linux - Restful web-services - JSON - Python	- RESTful API with HTTPS - SSL/TLS encryption for server communication - GDPR compliance	- Python 3.8 or later - Conda - PyTorch 1.8.0 or later

5 Conclusions & future work

In conclusion, the collection of software and hardware requirements for the system's components has provided a foundation for the development and integration of the SCENE system. By defining and gathering these specifications, we have ensured that all aspects of the project are thoroughly understood and accounted for. The ten fields for software requirements and seven fields for hardware requirements, coupled with detailed examples and descriptions, have facilitated a uniform and precise submission from all technical partners. This systematic approach has allowed us to capture specific requirements for each component, including those related to Generative AI, ensuring a robust and coherent set of project specifications.

The importance of collecting detailed software and hardware specifications cannot be overstated, as it is crucial for the successful implementation and operation of the system. These specifications serve as the blueprint for development, guiding the design, and ensuring compatibility and performance standards are met. Additionally, the first round of the state-of-the-art (SOTA) analysis is essential in identifying current technological capabilities and gaps, providing a benchmark for our system's performance. This initial analysis sets the stage for continuous improvement and innovation, ensuring that the SCENE system remains at the forefront of technological advancements and meets the evolving needs of its users.

References

[1] K. Hyeyoung, L. Suyeon, P. Yoonseo and C. Anna, "A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields," Electronic Solutions for Artificial Intelligence Healthcare, vol. II, 2022.

[2] T. Yi, A. T. Luu and H. C. Siu, "Latent Relational Metric Learning via Memory-based Attention for Collaborative Ranking," in WebConf: The Web Conference, Lyon France, 2018.

[3] B. Shilong, X. Qianqian, Y. Zhiyong, H. Yuan, C. Xiaochun and H. Qingming, "The Minority Matters: A Diversity-Promoting Collaborative Metric Learning Algorithm," in Neural Information Processing Systems, 2022.

[4] L. Youhua, D. Hanwen, N. Yongxin, Z. Pengpeng, G. Qi, Y. Fajie and Z. Xiaofang, "Multi-Modality is All You Need for Transferable Recommender Systems," in RecSys: ACM Conference On Recommender Systems, 2023.

[5] Y. Zheng, Y. Faije, S. Yu, L. Youhua, F. Junchen, Y. Fei, P. Yunzhu and N. Yongxin, "Where to Go Next for Recommender Systems? ID- vs. Modality-based Recommender Models Revisited," in IR: Research and Development in Information Retrieval, New York, 2023.

[6] [Online]. Available: http://www.purpleforge.com/index.php/knowledge-base/249-end-user-license-agreement-eula-for-a-mobile-app.

[7] European Comission, Regulation on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), 2012. Arndt, R., Troncy, R., Staab, S., & Hardman, L. (2009). Comm: A core ontology for multimediaannotation. *Handbook on Ontologies*, 403-421.

[8] Arndt, R., Troncy, R., Staab, S., & Hardman, L. (2009). Comm: A core ontology for multimediaannotation. Handbook on Ontologies, 403-421.

[9] Hunter, J. (2005). Adding Multimedia to the SemanticWeb: Building and Applying an MPEG-7 Ontology. Multimedia content and the semantic web: Methods, standards and tools, 75-106.

[10] Dasiopoulou, S., Tzouvaras, V., Kompatsiaris, I., & Strintzis, M. G. (2010). Enquiring MPEG-7 based multimedia ontologies. Multimedia tools and applications, 46, 331-370.

[11] Borgo, S., Ferrario, R., Gangemi, A., Guarino, N., Masolo, C., Porello, D., ... & Vieu, L. (2022). DOLCE: A descriptive ontology for linguistic and cognitive engineering. Applied ontology, 17(1), 45-69.

[12] Jin, H., Li, Y., Luan, F., Xiangli, Y., Bi, S., Zhang, K., ... & Snavely, N. (2024). Neural Gaffer: Relighting Any Object via Diffusion. arXiv preprint arXiv:2406.07520.

[13] Papantoniou, F. P., Lattas, A., Moschoglou, S., & Zafeiriou, S. (2023). Relightify: Relightable 3d faces from a single image via diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 8806-8817).

[14] Kocsis, P., Philip, J., Sunkavalli, K., Nießner, M., & Hold-Geoffroy, Y. (2024). Lightit: Illumination modeling and control for diffusion models. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 9359-9369).

[15] Zhao, X., Srinivasan, P. P., Verbin, D., Park, K., Brualla, R. M., & Henzler, P. (2024). IllumiNeRF: 3D Relighting without Inverse Rendering. arXiv preprint arXiv:2406.06527.

Funded by
the European Union

[16]Kim, H., Jang, M., Yoon, W., Lee, J., Na, D., & Woo, S. (2024). SwitchLight: Co-design of Physics-driven Architecture and Pre-training Framework for Human Portrait Relighting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 25096-25106).

[17]Fu, X., Yin, W., Hu, M., Wang, K., Ma, Y., Tan, P., ... & Long, X. (2024). Geowizard: Unleashing the diffusion priors for 3d geometry estimation from a single image. arXiv preprint arXiv:2403.12013.

[18]Johnson, R., & Zhang, T. (2014). Effective use of word order for text categorization with convolutional neural networks. arXiv preprint arXiv:1412.1058

[19]Tang, D., Qin, B., & Liu, T. (2015, September). Document modeling with gated recurrent neural network for sentiment classification. In Proceedings of the 2015 conference on empirical methods in natural language processing (pp. 1422-1432).

[20]Xu, J., Chen, D., Qiu, X., & Huang, X. (2016). Cached long short-term memory neural networks for document-level sentiment classification. arXiv preprint arXiv:1610.04989

[21]Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016, June). Hierarchical attention networks for document classification. In Proceedings of the 2016 conference of the North American section of the association for computational linguistics: human language technologies (pp. 1480-1489

[22]Zhou, X., Wan, X., & Xiao, J. (2016, November). Attention-based LSTM network for cross-lingual sentiment classification. In Proceedings of the 2016 conference on empirical methods in natural language processing (pp. 247-256).

[23]Dong, L., Wei, F., Tan, C., Tang, D., Zhou, M., & Xu, K. (2014, June). Adaptive recursive neural network for target-dependent twitter sentiment classification. In Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers) (pp. 49-54).

[24]Vo, D. T., & Zhang, Y. (2015, June). Target-dependent twitter sentiment classification with rich automatic features. In Twenty-fourth international joint conference on artificial intelligence.

[25]Wang, Y., Huang, M., Zhu, X., & Zhao, L. (2016, November). Attention-based LSTM for aspect-level sentiment classification. In Proceedings of the 2016 conference on empirical methods in natural language processing (pp. 606-615).

[26]Yang, M., Tu, W., Wang, J., Xu, F., & Chen, X. (2017, February). Attention based LSTM for target dependent sentiment classification. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 31, No. 1).

[27]J. Liu και Y. Zhang, «Liu, Jiangming ; Zhang, Yue,» in Proceedings of the 15th Conference of the European Section of the Association for Computational Linguistics: Volume 2, Short Papers, 2017.

[28]Sun, Z., Fan, C., Han, Q., Sun, X., Meng, Y., Wu, F., & Li, J. (2020). Self-explaining structures improve nlp models. arXiv preprint arXiv:2012.01786.

[29]Goodwin, T. R., Savery, M. E., & Demner-Fushman, D. (n.d.). Flight of the PEGASUS? Comparing Transformers on Few-Shot and Zero-Shot Multi-document Abstractive Summarization

[30]Xu, J., Chen, D., Qiu, X., & Huang, X. (2016). Cached long short-term memory neural networks for document-level sentiment classification. arXiv preprint arXiv:1610.04989.

[31]Liu, Y., & Lapata, M. (2020). Hierarchical transformers for multi-document summarization. ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference, 5070–5081. https://doi.org/10.18653/v1/p19-1500

[32]Zhang, J., Zhao, Y., Saleh, M., & Liu, P. J. (2019). PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. ArXiv.

[33]Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. http://arxiv.org/abs/1910.13461

[34]Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2019). Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. 1–53. http://arxiv.org/abs/1910.10683

[35]Jin, H., Wang, T., & Wan, X. (2020). Multi-Granularity Interaction Network for Extractive and Abstractive. Acl2020, 6244–6254. https://www.aclweb.org/anthology/2020.acl-main.556.pdf

[36]Goyal, T., Li, J. J., & Durrett, G. (2022). News summarization and evaluation in the era of gpt-3. arXiv preprint arXiv:2209.12356.

[37]Guo, S., Leborgne, F., Hu, J., & Baets, W. (2024). Can AI Bridge the Literacy Gap? Developing a GPT-4 Summarization Tool for Low Literacy. From User to Human, 52.

[38]Gu, J., Wang, G., Cai, J., & Chen, T. (2017). An empirical study of language cnn for image captioning. In Proceedings of the IEEE international conference on computer vision (pp. 1222-1231).

[39] Atliha, V., & Šešok, D. (2020, April). Comparison of VGG and ResNet used as Encoders for Image Captioning. In 2020 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream) (pp. 1-4). IEEE.

[40] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., ... & Bengio, Y. (2015, June). Show, attend and tell: Neural image caption generation with visual attention. In International conference on machine learning (pp. 2048-2057). PMLR.

[41] Yin, H., Vahdat, A., Alvarez, J. M., Mallya, A., Kautz, J., & Molchanov, P. (2022). A-vit: Adaptive tokens for efficient vision transformer. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10809-10818).

[42] Gan, Z., Li, L., Li, C., Wang, L., Liu, Z., & Gao, J. (2022). Vision-language pre-training: Basics, recent advances, and future trends. Foundations and Trends® in Computer Graphics and Vision, 14(3–4), 163-352.

[43] Lu, J., Batra, D., Parikh, D., & Lee, S. (2019). Vilbert: Pretraining task-agnostic visiolinguistic representations for vision-and-language tasks. Advances in neural information processing systems, 32.

[44] Barraco, M., Cornia, M., Cascianelli, S., Baraldi, L., & Cucchiara, R. (2022). The unreasonable effectiveness of CLIP features for image captioning: an experimental analysis. In proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4662-4670).

[45] Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. Government information quarterly, 36(2), 358-367.

[46] Petriv, Y., Erlenheim, R., Tsap, V., Pappel, I., & Draheim, D. (2020). Designing effective chatbot solutions for the public sector: A case study from Ukraine. In Electronic Governance and Open Society: Challenges in Eurasia: 6th International Conference, EGOSE 2019, St. Petersburg, Russia, November 13–14, 2019, Proceedings 6 (pp. 320-335). Springer International Publishing.

[47] Nirala, K. K., Singh, N. K., & Purani, V. S. (2022). A survey on providing customer and public administration based services using AI: chatbot. Multimedia Tools and Applications, 81(16), 22215-22246.

[48] Maher, S., Kayte, S., & Nimbhore, S. (2020). Chatbots & its techniques using AI: an review. International journal for research in applied science and engineering technology, 8(12), 503-508.

[49] Luo, B., Lau, R. Y., Li, C., & Si, Y. W. (2022). A critical review of state-of-the-art chatbot designs and applications. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 12(1), e1434.

[50] Nithuna, S., & Laseena, C. A. (2020, July). Review on implementation techniques of chatbot. In 2020 International Conference on Communication and Signal Processing (ICCSP) (pp. 0157-0161). IEEE.

[51] Luo, B., Lau, R. Y., Li, C., & Si, Y. W. (2022). A critical review of state-of-the-art chatbot designs and applications. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 12(1), e1434.

[52] Catapang, J. K., Solano, G. A., & Oco, N. (2020, February). A bilingual chatbot using support vector classifier on an automatic corpus engine dataset. In 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 187-192). IEEE.

[53] Rizou, S., Theofilatos, A., Paflioti, A., Pissari, E., Varlamis, I., Sarigiannidis, G., & Chatzisavvas, K. C. (2023). Efficient intent classification and entity recognition for university administrative services employing deep learning models. Intelligent Systems with Applications, 19, 200247.

[54] Manik, L. P., Akbar, Z., Mustika, H. F., Indrawati, A., Rini, D. S., Fefirenta, A. D., & Djarwaningsih, T. (2021). Out-of-Scope Intent Detection on A Knowledge-Based Chatbot. International Journal of Intelligent Engineering & Systems, 14(5).

[55] Stepin, I., Alonso, J. M., Catala, A., & Pereira-Fariña, M. (2020, July). Generation and evaluation of factual and counterfactual explanations for decision trees and fuzzy rule-based classifiers. In 2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 1-8). IEEE.

[56] Maeng, W., & Lee, J. (2021, May). Designing a chatbot for survivors of sexual violence: Exploratory study for hybrid approach combining rule-based chatbot and ml-based chatbot. In Proceedings of the Asian CHI Symposium 2021 (pp. 160-166).

[57] Bhanupriya, M., Kirubakaran, N., & Jegadeeshwari, P. (2023, December). EmotionTracker: Real-time Facial Emotion Detection with OpenCV and DeepFace. In 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI) (pp. 1-4). IEEE.

[58] Snavely, Noah, Steven M. Seitz, and Richard Szeliski. "Modeling the world from internet photo collections." International journal of computer vision 80 (2008): 189-210.

[59] Schonberger, J. L., & Frahm, J. M. (2016). Structure-from-Motion Revisited. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[60] Levoy, M., & Hanrahan, P. (1996). Light Field Rendering. In Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH).

[61] Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." Communications of the ACM 65.1 (2021): 99-106.

[62] Yu, Alex, et al. "Plenoctrees for real-time rendering of neural radiance fields." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

[63] Barron, Jonathan T., et al. "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields." Proceedings of the IEEE/CVF international conference on computer vision. 2021.

Funded by
the European Union

[64] Barron, Jonathan T., et al. "Zip-nerf: Anti-aliased grid-based neural radiance fields." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.

[65] Zwicker, Matthias, et al. "Surface splatting." Proceedings of the 28th annual conference on Computer graphics and interactive techniques. 2001.

[66] Yan, Zhiwen, et al. "Multi-scale 3d gaussian splatting for anti-aliased rendering." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

[67] Rusu, R. B., & Cousins, S. (2011). 3D is here: Point Cloud Library (PCL). In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA).

[68] Zhou, Qian-Yi, Jaesik Park, and Vladlen Koltun. "Open3D: A modern library for 3D data processing." arXiv preprint arXiv:1801.09847 (2018).

[69] Teerapong Serisamran, "The analysis of film editing techniques in Thai best editing feature film "Bad Genius" International Journal of Management and Applied Science, ISSN: 2394-7926 Volume-5, Issue-3, Mar.-2019.