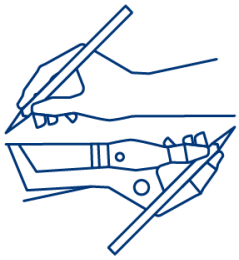
Funded by the European Union

Project: 101094364 — ITHACA — HORIZON-CL2-2022-DEMOCRACY-01

EUROPEAN RESEARCH EXECUTIVE AGENCY (REA)

REA.C – Future Society

C.1 – Inclusive Society



ITHACA

AI To Enhance Civic Participation

D3.3 Final ITHACA platform - report

Work Package 3:	ITHACA platform design and development
Authors:	KONNEKTABLE
Status:	Final
Due Date:	30/09/2025
Version:	1.0
Submission Date:	20/01/2026
Dissemination Level:	PU - Public

ITHACA Project Profile

Grant Agreement No.: 101094364

Acronym:	ITHACA
Title:	artificial Intelligence To enHAnce Civic pArticipation
URL:	https://www.ithaca-project.eu/
Start Date:	01/01/2023
Duration:	36 months

Partners

Short Name	Legal Name	Country
KT	**KONNEKT ABLE TECHNOLOGIES LIMITED**	**IE**
CERTH	**ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS**	**EL**
UPAT	**PANEPISTIMIO PATRON**	**EL**
RtF	**RAISING THE FLOOR**	**BE**
SnP	**STAMADIANOS KAI SYNETAIROI DIKIGORIKI ETAIREIA**	**EL**
UniGraz	**UNIVERSITAET GRAZ**	**AT**
MNLT	**MNLT INNOVATIONS IKE**	**EL**
SIMAVI	**SOFTWARE IMAGINATION & VISION SRL**	**RO**
PEDAL	**PEDAL CONSULTING SRO**	**SK**
BMA	**AGENTIA METROPOLITANA PENTRU DEZVOLTARE DURABILA BRASOV ASOCIATIA**	**RO**
	MESTO MARTIN	**SK**

Funded by the European Union

DOCUMENT HISTORY

VERSION	DATE	CHANGES	RESPONSIBLE PARTNER
0.1	04/09/2025	ToC	Epameinondas Koutavelis (KT)
0.2	30/09/2025	Feedback on the structure	Iliana Loi (UPAT), Adrian Dragota (SIMAVI), Katerina Touliou (CERTH), Michael Bedek (UniGraz), Evangelos Rigas (KT), Epameinondas Koutavelis (KT)
0.3	10/12/2025	Partners Contribution	Iliana Loi (UPAT), Adrian Dragota (SIMAVI), Laterina Toliou (CERTH), Michael Bedek (UniGraz), Evangelos Rigas (KT), Epameinondas Koutavelis (KT)
0.4	19/12/2025	Review and feedback	Evangelos Rigas (KT), Epameinondas Koutavelis (KT)
0.5	21/12/2025	Internal Review by partners	Katerina Touliou (CERTH), Adrian Dragota (SIMAVI)
1.0	20/01/2026	Finalisation	Evangelos Rigas (KT), Epameinondas Koutavelis (KT)

Table of Contents

1. EXECUTIVE SUMMARY

The **ITHACA** project (Artificial Intelligence To Enhance Civic Participation) is a pioneering initiative funded by **Horizon Europe**. Its core objective is to leverage **Artificial Intelligence (AI)** to enhance **civic engagement**, **democratic participation**, and **social inclusion**. The platform developed through ITHACA empowers citizens, especially those from **underrepresented and vulnerable groups**, to actively engage in **democratic processes** such as **policy-making**, **community discussions**, and **local governance**.

This deliverable, **D3.3 – Final ITHACA Platform Report**, represents a significant milestone in the project, documenting the evolution of the platform, particularly after insights gained from **pilot implementations** in **Brasov** (Romania) and **Martin** (Slovakia). The report details the latest **technical advancements**, **system architecture**, **AI tool integration**, and **user engagement** features of the platform.

Key achievements covered in **Work Package 3 (WP3)**, which this deliverable focuses on, include:

1. **Final Technical Requirements Review (T3.1)**:
    - The platform has been refined based on both **functional** (e.g., **evidence extraction**, **AI-based summarization**) and **non-functional** requirements (e.g., **performance**, **scalability**, **GDPR compliance**).
    - Updates have been made based on **real-world pilot feedback**, ensuring that the platform meets the evolving needs of both **citizens** and **local governments**.
2. **Final System Architecture (T3.2)**:
    - The platform's **hybrid architecture**, combining **microservices** with an **event-driven model**, ensures scalability, flexibility, and enhanced **security**.
    - Key **trust**, **security**, and **compliance mechanisms** have been integrated, in line with EU regulations, including **GDPR** and **ethical AI guidelines**.
3. **Core Components Implementation**:
    - Major components such as **evidence extraction**, **AI-driven cognitive agents**, **argument visualization**, and **gamification incentives** have been successfully developed and integrated into the platform.
    - These components aim to **enhance civic participation**, providing tools for **citizens** to **engage in discussions**, **submit proposals**, and **vote** on important civic issues.
4. **Platform Integration and Operation**:
    - The platform has been successfully integrated with **external systems** and deployed in **pilot cities**, offering **AI-powered tools** such as **toxicity detection** and **summarization** to manage citizen discussions.
    - Feedback from **pilot users** highlighted key strengths in usability but also pointed out areas for improvement, such as **AI transparency**, **gamification fairness**, and

accessibility.

5. **Testing and Validation**:

    - **Comprehensive testing** (including **performance**, **load**, **security**) and **user acceptance testing** have been carried out, leading to several **iterative improvements** that ensure the platform is stable, user-friendly, and meets required standards.

Aligned with **EU objectives for inclusive democracy**, the **ITHACA platform** sets a new precedent in **civic technology**, offering a **scalable solution** for local governments and civil society to facilitate **engagement**. Moving forward, the platform is positioned for **long-term sustainability**, with exploitation strategies that include **municipal partnerships**, **open-source community involvement**, and **potential revenue models**. Future development will focus on improving **AI fairness**, **platform usability**, and **expanding international reach** through **multilingual support** and **cross-platform integration**.

In conclusion, **D3.3** serves not only as a **technical report** detailing the final design, implementation, and validation of the **ITHACA platform**, but also as a **strategic document** that outlines its path towards empowering citizens and enhancing **democratic participation** across Europe.

2. Introduction

2.1 Background of the ITHACA Project

The **ITHACA project** (Artificial Intelligence To Enhance Civic Participation) is an EU-funded initiative under **Horizon Europe** aimed at developing innovative solutions to enhance **democratic engagement** through the integration of **Artificial Intelligence (AI)**. The project addresses the growing need for more **inclusive**, **transparent**, and **efficient** forms of civic participation, particularly for **vulnerable** and **underrepresented groups** in society.

The primary goal of ITHACA is to develop a platform that enables **citizens** to engage in democratic processes, such as **policy-making**, **community discussions**, and **local governance** in a more effective and meaningful way. By integrating **AI tools** such as **evidence extraction**, **toxicity detection**, **summarization**, and **argument visualization**, the platform enhances the quality of civic discourse, ensuring that **AI-driven solutions** complement human interaction without undermining the democratic process.

The project is carried out by a **consortium of partners** from across Europe, bringing together expertise in **AI**, **civic technologies**, **democratic processes**, and **ethics**. Through the use of AI technologies, ITHACA aims to support the creation of transparent, participatory, and effective mechanisms for public decision-making at the local and regional levels.

2.2 Purpose and scope of Deliverable 3.3

Deliverable D3.3 presents the final iteration of the **ITHACA platform**, building upon the initial designs, concepts, and technical implementations presented in **D3.1** and **D3.2**. This deliverable focuses on reviewing and finalizing the platform's **technical requirements**, **system architecture**, and **core components** following the **second evaluation phase**, which provided critical feedback from pilot implementations in **Brasov** (Romania) and **Martin** (Slovakia).

The purpose of **D3.3** is to document the final status of the platform, ensuring that it aligns with the core objectives of **enhanced civic participation**, **AI transparency**, and **system robustness**. The scope of this deliverable covers:

- A **review of technical requirements** based on the latest pilot evaluations, including **functional** and **non-functional** updates.
- A **final system architecture**, including key updates to the platform's **microservices** and **event-driven models** for **scalability** and **resilience**.
- Detailed information on the **core components** such as **AI-driven moderation**, **argument visualization**, and **gamification**, with insights from real-world testing.

- **Integration** and **operation strategies** based on the performance in pilot sites.

By presenting the latest version of the platform, this deliverable provides a roadmap for future developments and sets the foundation for the **sustainability** and **exploitation** of the platform beyond the scope of the project.

2.3 Relation to previous deliverables

This deliverable builds on the earlier reports **D3.1**:

- **D3.1**: The first version of the deliverable outlined the **initial technical requirements**, including **functional** aspects such as **user engagement features**, **AI tools**, and **core components**. It also described the platform's **architecture** and addressed initial challenges identified during the early stages of development.

In **D3.3**, the latest feedback from the **second evaluation phase** has been integrated, resulting in **finalized updates** to the technical requirements and system components. This report reflects the outcomes of the **pilot testing** and offers a **comprehensive review** of the platform's maturity and readiness for future use.

2.4 Structure of the document

This deliverable is structured to provide a clear and progressive overview of the final design, implementation, validation, and evaluation of the **ITHACA platform**, following the logical flow of the platform's development lifecycle and the objectives of **Work Package 3**.

- **Chapter 1 – Executive Summary**
  This chapter provides a concise overview of the ITHACA project and the scope of Deliverable D3.3. It highlights the main objectives, key technical achievements, and outcomes of the platform's final development phase, with particular emphasis on insights derived from pilot implementations in Brasov and Martin.

- **Chapter 2 – Introduction**
  This chapter sets the context for the deliverable by presenting the background of the ITHACA project, the purpose and scope of D3.3, and its relation to previous technical deliverables (D3.1 and D3.2). It also outlines the structure of the document to guide the reader through its contents.

- **Chapter 3 – Final Technical Requirements Review (T3.1)**
  This chapter presents the final review of the platform's technical requirements, including updates to functional and non-functional requirements based on pilot feedback and

evaluation results.

- **Chapter 4 – Final System Architecture (T3.2)**
  This chapter describes the final system architecture of the ITHACA platform, including the hybrid microservices and event-driven approach, deployment model, and the integration of security, trust, and compliance mechanisms.

- **Chapter 5 – Final Implementation of Core Components**
  This chapter details the implementation of the platform's core components, such as cognitive and social agents, gamification incentives, argument visualization, the Public AI Register, trust and security infrastructure, personal information management, and continuous integration and deployment pipelines.

- **Chapter 6 – Platform Integration and Operation**
  This chapter explains how the platform operates from the user's perspective and how it was deployed and used in real-world pilot contexts. It includes detailed user workflows and an overview of platform operation in Brasov and Martin.

- **Chapter 7 – Testing and Validation (T3.11)**
  This chapter presents the testing and validation activities carried out for the platform, including technical testing, user acceptance testing, and pilot testing results.

- **Chapter 8 – Challenges, Lessons Learned, and Mitigations**
  This chapter discusses the main technical, ethical, legal, usability, and accessibility challenges encountered during development and pilot deployment, along with the lessons learned and mitigation strategies applied.

- **Chapter 9 – Conclusions and Future Work**
  The final chapter summarizes the overall outcomes of the deliverable, assesses alignment with the Grant Agreement objectives, outlines sustainability and exploitation perspectives, and provides recommendations for future development of the ITHACA platform.

3. Final Technical Requirements Review (T3.1)

The **ITHACA platform** is a modular, AI-enabled digital infrastructure designed to support **inclusive, transparent, and trustworthy civic participation** at the local level. It enables citizens, moderators, and public authorities to engage in structured online deliberation through a set of integrated tools for discussion, evidence extraction, argument visualisation, and AI-assisted moderation, while ensuring compliance with European legal and ethical requirements.

The platform has been developed following a **user-centred and iterative approach**, combining insights from participatory design activities, pilot deployments, and evaluation phases carried out in real-world municipal contexts. Its technical design is grounded in a **hybrid architecture**, supporting scalability, interoperability, and resilience, and incorporates privacy-by-design and human-in-the-loop principles to ensure responsible use of Artificial Intelligence.

This chapter provides the final review of the **technical requirements** for the **ITHACA platform**, incorporating feedback from the **second evaluation phase** and pilot testing results. The goal is to ensure that the platform meets the evolving needs of **citizens**, **local governments**, and other stakeholders, while maintaining **scalability**, **performance**, **security**, and **compliance**. The review focuses on updates to both **functional** and **non-functional requirements** based on real-world testing and feedback from the pilot implementations in **Brasov** (Romania) and **Martin** (Slovakia).

3.1 Functional requirements – updates from pilots

As foreseen in the Grant Agreement, the definition and refinement of the platform's functional requirements followed an **iterative and evaluation-driven process**. While the initial planning envisaged the availability of the first consolidated version of the platform at the beginning of 2025, the consortium jointly agreed to **adjust the timeline** in order to incorporate the results of the **first evaluation phase conducted in the pilot sites**. This decision aimed to ensure that the technical requirements and platform functionalities were properly aligned with real user needs and operational conditions observed during the pilots.

Similarly, the final version of the platform was deliberately scheduled after the completion of all pilot activities, allowing the consortium to **integrate the full set of evaluation findings** and fine-tune the platform accordingly. As some pilot activities experienced minor delays, the finalisation of the platform was correspondingly shifted. This approach ensured that **D3.3 reflects a mature, validated, and evidence-based implementation**, rather than a purely theoretical or prematurely frozen technical solution.

The **functional requirements** of the ITHACA platform were revisited and refined based on the outcomes of the second evaluation phase, which included **pilot testing** in the cities of **Brasov** and **Martin**. These updates ensure that the platform's features are user-centric, efficient, and aligned with the goals of enhancing civic participation.

1. **Citizen Engagement Features**:

- **Discussions and Proposals**: One of the key features of the platform is to allow citizens to **create proposals** and participate in **discussions**. Based on pilot feedback, the proposal creation and voting processes were **streamlined** to ensure better usability. This includes clearer steps for submitting proposals and more **intuitive navigation** through ongoing discussions.

- **Argument Visualization**: The integration of **argument visualization** tools was widely appreciated, helping citizens to follow complex debates more effectively. However, users suggested enhancing the **clarity** of visual representations, which led to the inclusion of **interactive guides** that explain the visualized arguments in detail.

2. **AI-Based Moderation and Summarization**:

   - **AI Summarization**: The platform's AI-driven **summarization tool** was improved based on user feedback. The summarizations were found useful, but users emphasized the need for the tool to reflect a broader range of viewpoints, especially those of **minority groups**. As a result, the AI summarization tool was adjusted to better incorporate **underrepresented opinions**.

   - **Toxicity Detection**: The **AI moderation tool** was updated to improve its accuracy in detecting harmful content while minimizing false positives. Users requested more **transparent feedback** on why content was flagged, leading to the integration of explanations alongside flagged content.

3. **Gamification and Engagement**:

   - **Gamification Features**: To motivate participation, the platform introduced a **gamification system** incorporating **missions**, **badges**, and **leaderboards**. However, feedback indicated confusion around the **rules** for earning rewards, particularly for new users. The **point system** was revised to focus on **meaningful engagement** rather than sheer activity, and the **leaderboard** was adjusted to encourage more **inclusive participation**.

   - **User Engagement**: Engagement was measured through participation metrics and the **gamification system**, and it was noted that **continuous feedback** on user contributions could further increase **long-term participation**. Updates include **real-time feedback** on users' participation progress and **clarified reward criteria**.

3.2 Non-functional requirements (performance, scalability, accessibility, GDPR/AI Act compliance)

Non-functional requirements define the operational standards the platform must meet to function effectively under real-world conditions. This section provides updates on **performance**, **scalability**, **accessibility**, and **compliance** with **GDPR** and the **EU AI Act** based on the evaluation results.

1. **Performance and Scalability**:

    - **Challenge**: During the pilot phase, the platform experienced performance issues during periods of **high traffic**, particularly in relation to **data processing** and **user load**. While the system performed well under normal conditions, **peak usage** periods, such as public voting events, presented scalability challenges.

    - **Update**: To address these concerns, the platform's backend was enhanced to ensure better **auto-scaling** and **load balancing**. The cloud infrastructure was further optimized to handle fluctuating traffic, ensuring high availability and low latency. In addition to internal performance optimisation activities, **stress and load testing** were conducted to validate the platform's behaviour under increased user demand and peak usage scenarios. These tests were aligned with the **technical performance evaluation activities carried out during the pilot phase**, as reported in **Deliverable D4.3**, ensuring consistency between system-level testing and real-world operational conditions observed in the pilot deployments. The results confirmed that the platform meets the defined performance and scalability requirements for the expected usage scenarios in municipal contexts.

2. **Accessibility**:

    - **Challenge**: Although the platform's **accessibility features** were positively received, some users found it difficult to locate or activate certain **assistive tools**, particularly those related to **visual impairments** and **keyboard navigation**.

    - **Update**: Based on feedback, the **accessibility features** were made more visible and intuitive. A dedicated **accessibility menu** was introduced, and the platform's **user interface** was simplified to reduce cognitive overload, making it easier for users with varying abilities to navigate.

    - Additional improvements, such as **speech-to-text** functionality and enhanced **text-to-speech** capabilities, were also added to ensure the platform is **fully accessible** for users with disabilities.

3. **GDPR and AI Act Compliance**:

   - **Challenge**: Ensuring compliance with **GDPR** (General Data Protection Regulation) and the **EU AI Act** was a major consideration during the development of the platform. Managing **personal data**, ensuring **user consent**, and providing **AI transparency** were areas of focus.

   - **Update**: The platform's compliance mechanisms were strengthened to ensure **GDPR compliance**. The **Personal Information Management System (PIMS)** was refined, allowing users to **manage** and **control** their data effectively. **Anonymization** protocols were enhanced to safeguard user privacy, and **data retention policies** were clearly communicated to users.

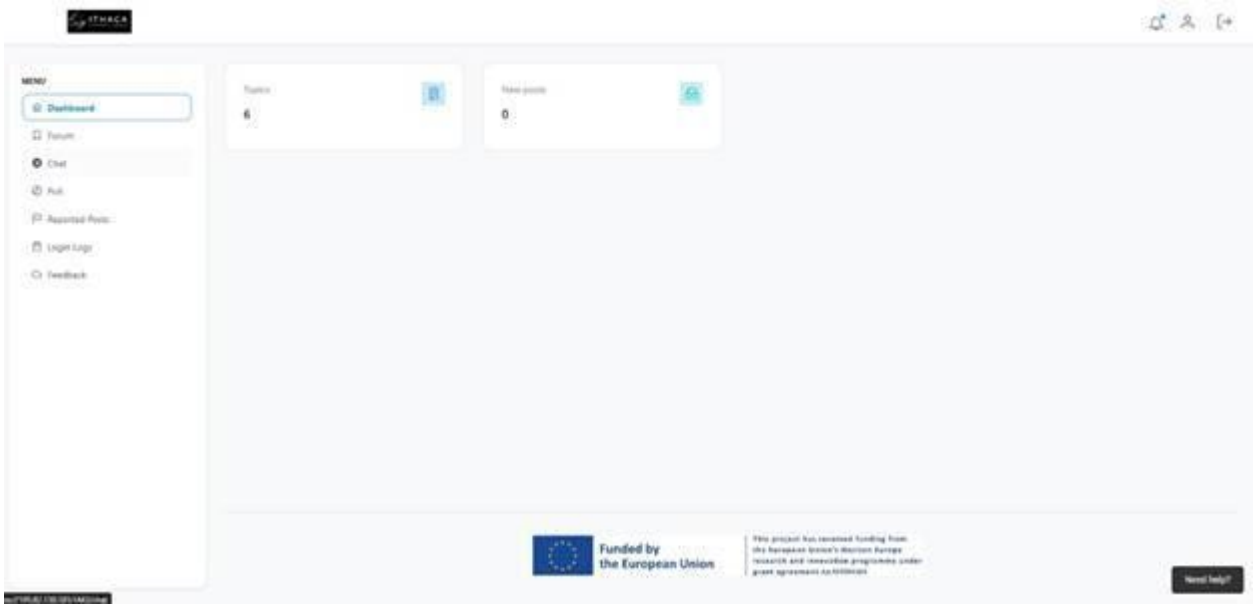
   - The platform's **AI tools** underwent an evaluation process to ensure that they align with the **EU AI Act**, focusing on **fairness**, **accountability**, and **non-discrimination**. Transparency features were introduced, such as **AI decision explanations** and **audit trails**, allowing users to understand how AI-driven tools make decisions.

3.3 Adjustments based on evaluation feedback

The second evaluation phase provided valuable insights that led to further refinement of the platform. These adjustments were made to ensure the platform meets the expectations of users, stakeholders, and regulatory bodies:

1. **Refinement of AI Tools**:
    - Based on user feedback, the **AI moderation system** was updated to reduce the incidence of **false positives** and improve its ability to detect and appropriately flag harmful content. The **summarization algorithms** were also fine-tuned to ensure more **balanced** content representation, incorporating a wider variety of perspectives.
2. **Improved User Interface (UI)**:
    - Feedback from **pilot users** highlighted areas of the **UI** that were challenging to navigate. The **dashboard** was simplified, with **clearer visual cues** for key actions like posting content, submitting proposals, and voting. The **onboarding process** was also improved with interactive **tutorials** that guide users through the platform's features step by step.

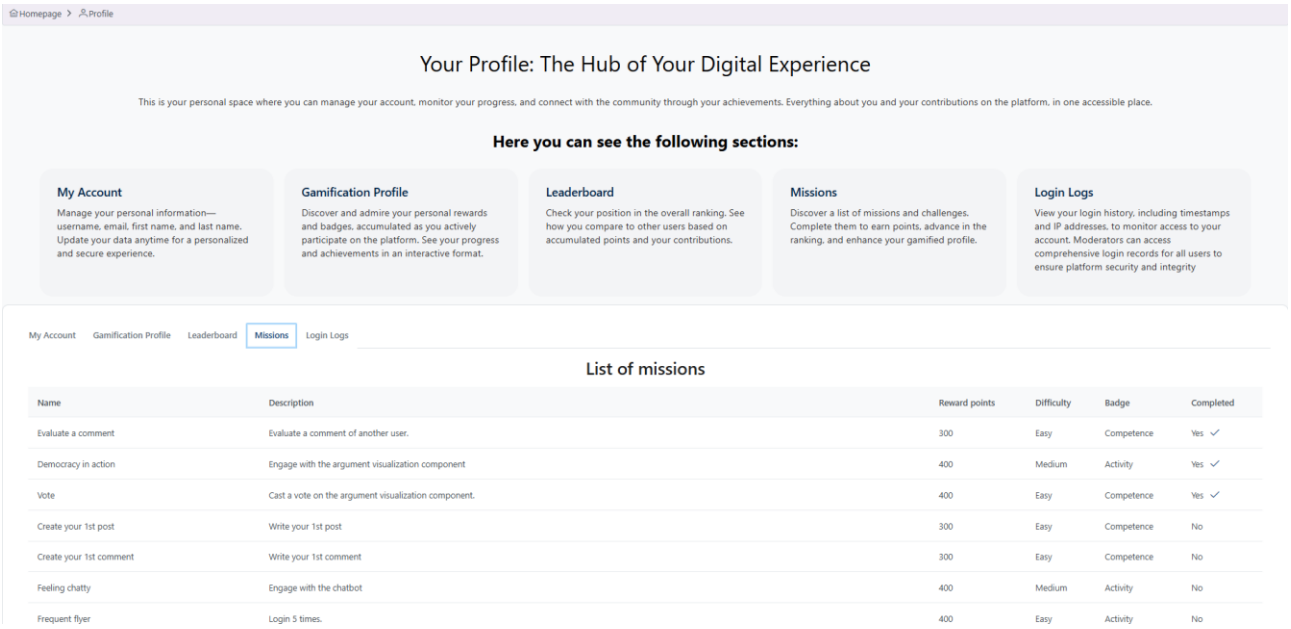
Old landing page:



New landing page:

3. **Enhanced Gamification System**:

    - Several users reported that the **leaderboard** system could be demotivating for those who did not regularly participate. To address this, the **gamification features** were adjusted to reward **meaningful contributions**, such as **engagement in debates** or **constructive feedback**, rather than just activity. The reward system was made more **inclusive**, with **points** for users contributing in different ways.



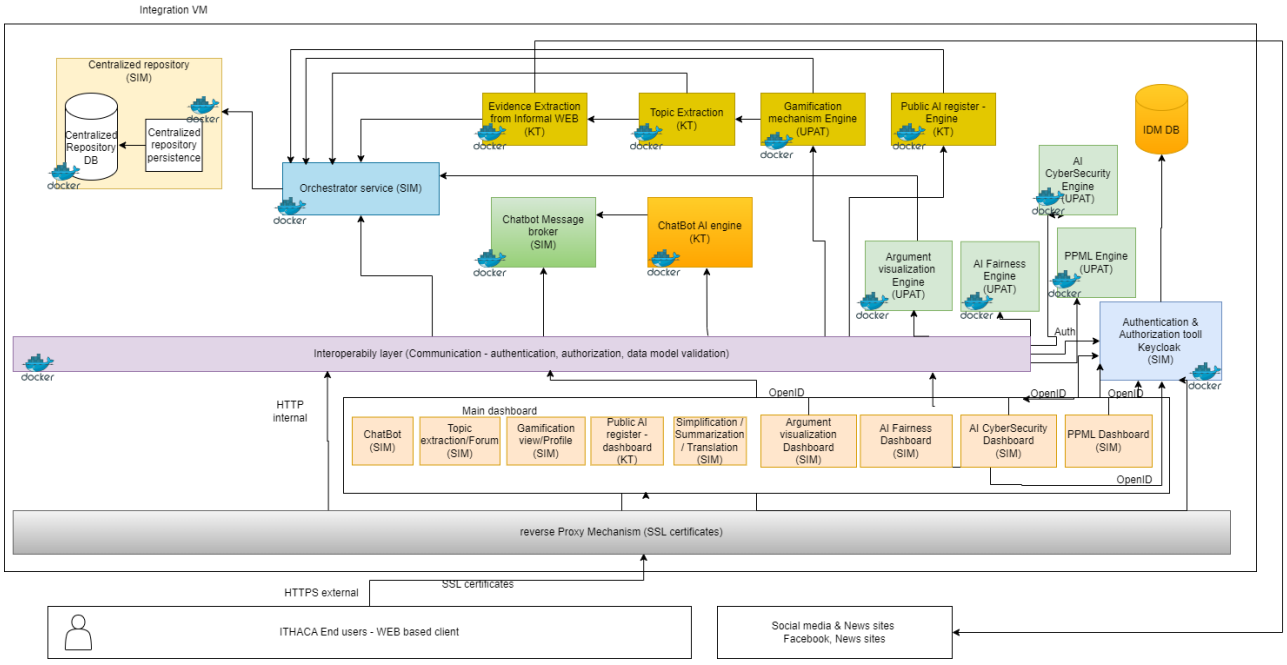
4. **Integration with External Systems**:

    - In response to user feedback, the platform's ability to integrate with **external civic systems** (such as local government platforms and third-party data sources) was enhanced. This ensures that the platform can be **interoperable** with existing **public sector tools**, enabling seamless data exchange and enhancing the platform's value in real-world civic contexts.

The updates made to the **functional** and **non-functional requirements** ensure that the **ITHACA platform** is equipped to meet the needs of a diverse set of users, ranging from citizens to government officials. Based on **pilot testing** and real-world feedback, the platform's **scalability**, **usability**, **AI fairness**, and **compliance** have been optimized, paving the way for future development and deployment. These refinements are key to ensuring that ITHACA can be deployed at scale, helping foster greater **democratic engagement** and **civic participation** across Europe.

4. Final System Architecture (T3.2)

In this chapter, we provide a comprehensive overview of the **final system architecture** of the **ITHACA platform**, based on the insights gained through the **second evaluation phase** and **pilot testing**. The architecture outlines how the platform integrates various **components**, including its **hybrid microservice architecture**, **AI tools**, **security mechanisms**, and **deployment model**. This chapter also explores how the system is designed for scalability, flexibility, and compliance with European standards.



4.1 Overview of hybrid architecture (microservices + event-driven)

The **ITHACA platform** uses a **hybrid architecture** that combines the best of both **microservices** and **event-driven** models. This architecture was designed to ensure that the platform is **scalable**, **resilient**, and capable of handling the high volume of user interactions that come with democratic engagement tools.

1. **Microservices Architecture**:

    - **Modular Design**: The platform is broken down into discrete **microservices**, each responsible for a specific function, such as **user authentication**, **AI summarization**, **moderation**, **proposal management**, and **gamification**. This design enables **independent scaling** of each component and reduces the risk of system failures affecting the entire platform.

    - **Loose Coupling**: Each microservice communicates with others via **APIs** or **message queues**, ensuring that individual services can be updated, deployed, and

scaled without disrupting the overall system.

- **Technological Stack**: The backend is built using **Spring Boot** for Java-based services, while **Angular** is used for the **frontend**. Microservices interact with a central database and distributed systems that ensure high availability and fault tolerance.

2. **Event-Driven Architecture**:

   - The system incorporates an **event-driven model** to handle **real-time user interactions** and ensure that updates (such as proposal voting or comment moderation) are propagated efficiently across the platform.

   - **Event Bus**: An **event bus** handles **asynchronous communication** between services, ensuring that actions like **user posts**, **AI analysis**, and **content updates** are processed independently without slowing down the platform's responsiveness.

   - **Decoupling of Services**: By using events to trigger system actions, the platform ensures that services remain **decoupled**, allowing for easier maintenance and future upgrades.

3. **Benefits of the Hybrid Model**:

   - **Scalability**: The system can handle increased demand by independently scaling individual microservices or adding additional event-processing resources.

   - **Resilience**: The event-driven model ensures that failure in one service does not bring down the entire platform, allowing for continuous operation even in the face of errors or unexpected loads.

   - **Flexibility**: This architecture supports a variety of civic engagement use cases and allows for the future addition of new services or integration with third-party systems.

4.2 Deployment model (cloud, scalability, redundancy)

The **deployment model** of the **ITHACA platform** has been designed to maximize its **scalability**, **availability**, and **fault tolerance**. It is built on cloud infrastructure, ensuring that the platform can dynamically adjust to the number of active users and the demands of different environments (e.g., pilot cities or future deployments).

1. **Cloud-Based Infrastructure**:

   - The platform is hosted on **cloud infrastructure** to take advantage of **on-demand scalability** and **cost-effective resource management**. This deployment model

ensures that the platform can be easily scaled to accommodate growing user bases and high traffic periods during **civic events**.

- The **cloud** also enables the use of **distributed services**, which ensures that data is **replicated** across multiple servers, improving system availability and reducing latency.

2. **Scalability and Auto-Scaling**:

   - The platform is designed to scale automatically based on **user activity**. For example, during peak times (such as civic engagement campaigns or voting periods), the platform can automatically allocate additional resources to **handle increased load**.

   - **Horizontal scaling** is used, where additional instances of microservices are launched to balance the workload across multiple machines or cloud containers.

3. **Redundancy and High Availability**:

   - **Redundancy** is built into the system to ensure **continuous availability**. Multiple instances of each **microservice** and critical components (e.g., databases, application servers) are deployed across different geographic locations, ensuring that the platform remains operational even in the event of server failures or regional outages.

   - **Load Balancers** ensure that user requests are distributed evenly across available resources, improving system performance and minimizing downtime.

4. **Disaster Recovery**:

   - **Backup systems** are implemented to maintain regular data snapshots, ensuring that the platform can recover quickly in the event of a **data corruption** or **system failure**. These backups are stored in geographically **distributed locations** to ensure resilience against regional failures.

4.3 Security, trust and compliance mechanisms

Security, trust, and compliance are foundational to the success of the **ITHACA platform**, given its role in handling **citizen participation** and **sensitive data**. The platform is designed to meet high standards of security, adhering to **GDPR** and other relevant **EU regulations**.

1. **Authentication and Authorization**:

- The platform uses **Keycloak** as the central **Identity and Access Management (IAM)** solution, ensuring that all users (citizens, moderators, admins) are securely authenticated and authorized to access appropriate services.

- **Single Sign-On (SSO)**: Keycloak supports **SSO**, enabling users to log in once and access all platform services without repeated logins.

- **Role-Based Access Control (RBAC)**: Different levels of access are granted based on the user's role (e.g., **Citizen**, **Moderator**, **Admin**). This ensures that sensitive actions, such as moderating content or managing AI models, are restricted to authorized users only.

2. **Data Security**:[1]

   - **Encryption**: All sensitive data, including **user personal information**, is encrypted both **in transit** (via **TLS: Transport Layer Security/SSL: Secure Sockets Layer**) and **at rest** (via **AES: Advanced Encryption Standard encryption**). This ensures that even in the event of a security breach, the data remains protected.

   - **Anonymization**: The platform implements **data anonymization** techniques to ensure that personal data is not exposed during AI analysis or in user feedback reports.

   - **Secure APIs**: All internal and external APIs are secured using **OAuth 2.0** and **JWT (JSON Web Tokens)** to ensure that only authorized services can interact with the platform's components.

3. **Trust and Transparency**:

   - The **Trust & Security Infrastructure** includes an AI transparency dashboard, where users can view how the platform's **AI models** operate, the types of data they use, and how decisions (such as content moderation or proposal summaries) are made. This increases **user trust** in the platform's AI tools.

   - The platform also integrates a **feedback mechanism**, allowing users to challenge AI decisions, report issues, and receive transparency about AI actions.

4. **Compliance with GDPR and EU AI Act**:[2]

   - The platform is fully **GDPR-compliant**, ensuring that **personal data** is handled securely and transparently. Users can control their data and withdraw consent at any time.

[1] https://gdpr-info.eu/art-32-gdpr/

[2] https://ai-act-law.eu/?utm_source=gdpr-info.eu

- **AI ethics** are a core component of the platform, and regular audits ensure that all AI models comply with the **EU AI Act** [3]in terms of **fairness**, **non-discrimination**, and **transparency**.

4.4 Integration of components

The **ITHACA platform** integrates various core components, each contributing to the overall functionality of the system. These components were developed and refined based on the feedback from the second evaluation phase and are designed to operate seamlessly together.

- **Evidence Extraction from Citizen Discussions**:
  - The **Evidence Extraction Module** processes citizen discussions to extract relevant data and categorize topics. It uses **Natural Language Processing (NLP)** techniques to identify key themes and generate summaries.
  - This module is integrated with the **AI moderation system** to flag inappropriate content, ensuring that discussions remain civil and aligned with community guidelines.
- **Cognitive and Social Agents**:
  - The **Cognitive and Social Agents** interact with users and assist in moderating discussions, providing suggestions, and answering questions. These agents are designed to simulate human-like interactions while maintaining transparency and accountability.
- **Gamification Incentives**:
  - The **gamification system** rewards users for their participation and engagement on the platform. It includes points, badges, and leaderboards to encourage active involvement in civic discussions.
- **Public AI Register**:
  - The **Public AI Register** provides transparency about the AI tools used on the platform, listing the algorithms, their purposes, and how they are monitored for fairness and accountability.
- **Argument Visualization**:

[3] https://ai-act-law.eu/?utm_source=gdpr-info.eu

- This module visualizes citizen-generated arguments, helping users navigate complex discussions and identify key points of contention. The **visualization** supports real-time updates and enables users to engage with the arguments dynamically.

As the platform evolves, this architecture provides the foundation for **continuous innovation** and **future expansion**, enabling ITHACA to serve as a model for **inclusive democracy** and **AI-enhanced civic engagement**.

5. Final Implementation of Core Components

5.1 Cognitive and Social Agents

This task involves the design of Natural Language Processing (NLP)-based components for clustering and identifying discussion topics, as well as developing automated content moderation mechanisms to assist human-based moderation. These components aim to contribute to a deeper understanding of discussions and improve the safety and relevance of platform interactions. Texts to be processed include topics created on the platform, along with web content collected via the tools described in Chapter 4 (T3.3).

5.1.1 Categorization and clustering of texts

Initially, two methods were explored:

5.1.1.1 A. Categorization through sequence classification

This method first translates the Romanian and Slovak texts using the translation models described in 4.3.2 and subsequently passes the resulting English texts as tokens (with the aid of transformers.AutoTokenizer) to the *cardiffnlp/tweet-topic-large-multilingual* sequence classification model. From there, topic labels are predicted for each text. This step was later refined to use embeddings extracted with the paraphrase-multilingual-MiniLM-L12-v2 model, along with similarity thresholds, in order to build collections of topics and assign them to textual resources.

5.1.1.2 B.Clustering and keywords through a sentence transformer

This method involves using a sentence transformer, the model that has been tested is paraphrase-multilingual-MiniLM-L12-v2 - and then passing the embeddings created to a clustering algorithm (K-Means module was selected for that). The final output is a collection of clusters along with their keywords.

The main caveats of this process which were researched and addressed are:

- Finding out if it is more reliable to first translate the original content or pass it to the sentence transformer after some Unicode normalization (due to special characters existing in Romanian and Slovak)
- Identifying the appropriate number of topics for clusters
- Using stop words for the vectorizer that identifies keywords
- Assessing the keywords' value (e.g., some keywords may be irrelevant or just common words)

The role of the vocabulary dataset in categorization and clustering

As the project evolved and the vocabulary dataset was developed and refined (see 5.1.3), so did the topic extraction and categorization. With the use of the vocabulary, embeddings could act as comparison targets and topic assignment was refined to be informed by the vocabulary about what topics exist. Instead of just discovering topics on the fly, the vocabulary assists in mapping text to existing concepts and topic normalization, while also trying to mitigate the problem of language differences, thanks to the anchoring of concepts that happens in the vocabulary dataset development phase.

Subsequently, the clustering process benefits from the development of the vocabulary dataset, by vectorizing concepts and producing a semantic fingerprint. Top vocabulary concepts become cluster labels with the help of clustering utilities inside the *sklearn (scikit-learn)* module. This step also helps reduce the multilingual challenges presented, while also offering explainability patterns regarding the clustering process, in order to be better informed about how to improve these procedures.

5.1.2 Automated moderation

The core part of this task is the development of an automated moderator ("bot"), which will ensure that content uploaded to the platform meets community guidelines for safety and appropriateness. All usersubmitted posts and comments are submitted to classification in order to detect (among other issues to be considered) hate speech, toxic language and harassment or offensive content. This classification extends to uploaded images as well, which are checked for inappropriate or harmful content using image classifiers.

Moderation workflow is the following:

● Analysis of submitted text and images through an API. At this point, content is uploaded but not yet visible to the platform, as it is pending validation.

● Automatic rejection, flagging, or approval of content based on classifier outputs

The classifier used for text is unitary/toxic-bert, whose behaviour seems the most consistent so far, with other candidates being tested: Hate-speech-CNERG/dehatebert-mono-english and google/shieldgemma-2b. Text goes through the selected model's text-classification pipeline and outputs probabilities of inappropriate content.

For image moderation, the CLIP (dual-encoder model trained on image-text pairs) model openai/clipvit-base-patch32 has been selected to evaluate potential inappropriate content. This model produced better results so far than other candidates, the latter being: google/vit-base-patch16-224, google/vitlarge-patch16-384 and openai/clip-vit-large-patch14. The evaluation process involves resizing the image, setting labels of inappropriate content and setting a probability threshold between 0 and 1. The model then classifies the image, assigning a probability per label and if any label surpasses the threshold, the image is deemed inappropriate. In order to extract more reliable results from the image classification tool, the solution of fine-tuning the model is being investigated. This will involve using a dataset of images with toxic category labels. The dataset considered for this process is the LAION NSFW subset (laion/laion2B-en-nsfw).

5.1.3 Vocabulary Dataset

A dedicated vocabulary dataset has been developed from scratch as part of this task. The purpose of this resource is to identify key terms, phrases, and expressions used by citizens and public authorities in the context of e-government, city services, and local issues. This helps improve the functionalities of clustering, keyword extraction and content tagging. This process would involve gathering a representative collection of texts to mine vocabulary from the platform posts and the content retrieved from the web, as described in Chapter 4. An extra step of human involvement for filtering the resulting vocabulary and fine-tuning was taken.

The first objective towards achieving this was to create an initial project-specific vocabulary dataset reflecting issues regarding civic life and participation, such as municipal administration topics, public service announcements, infrastructure and events, in order to bootstrap the whole process. The source material consisted of a selected pool of articles inside the official municipal webpages of

Martin and Brasov, obtained and parsed with the help of the web crawling tool that was developed (see D3.4, section 4.3.4). The process was designed to work with limited initial data, support multilingual content and produce outputs suitable for semantic modeling.

After this step, keyphrase extraction was performed using KeyBERT, a semantic keyphrase extraction library that identifies words and phrases that are semantically representative of the whole document, using embeddings from BERT models (Bidirectional Encoder Representations from Transformers). KeyBERT does not rely purely on term frequency and works well on small text collections, which are the reasons it was chosen. KeyBERT was configured with the multilingual sentence embedding model sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2, which was suitable for supporting the Romanian and Slovak languages. The output of this stage was a list of candidate relevant keyphrases per document. Candidate terms were then translated to English, which was used as a pivot language, while also retaining the original wording. This step helps to reliably detect that two terms in different languages refer to the same underlying concept, especially in the case of comparing Slovak with Romanian, where high-quality bilingual resources and similarity models are scarce.

Afterwards, a refinement phase took place, which consolidated lexical and semantic variants, assigned provisional concept categories, and enriched entries with stable identifiers and metadata, resulting in the ontology-ready dataset. This phase involved lexical normalization, canonical form selection and semantic similarity consolidation and clustering, using embeddings from the sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2 model. The product of this process is a concept-level vocabulary, which was subsequently RDF / SKOS serialised using RDFLib and Turtle, ensuring interoperability and readiness for ontology alignment and semantic indexing.

5.1.4 Sentiment Analysis

5.1.4.1 Overview

Sentiment analysis is conceived as a mechanism for understanding citizens' attitudes, concerns, and perceptions regarding public services, municipal initiatives, and topics of civic relevance. The objective of this task is not limited to producing quantitative sentiment scores, but rather to support qualitative insights that can inform deliberation processes, policy design, and institutional responsiveness. The sentiment analysis framework is designed to operate on multilingual content in Romanian and Slovak.

5.1.4.2 Constraints

Social Media Data Sources

Although public social media platforms constitute a valuable source of civic discourse, performing sentiment analysis on such data was subject to external technical, legal, and ethical constraints.

In the case of Facebook, access to post and comment data is governed by a restrictive API policy that significantly limits automated data collection, particularly for research-oriented processing that involves large-scale text analysis. Furthermore, considerations related to user privacy and responsible data usage require a cautious approach when handling social media content, even when such content is publicly accessible.

Similarly, access to data from platforms such as Twitter (X) is constrained by pricing models that are not proportionate to the analytical needs of the project, making systematic sentiment analysis economically unfeasible within the current framework.

As a result, no sentiment analysis has been performed on social media content during this phase. This decision reflects a deliberate prioritization of legal compliance, ethical data handling, and sustainability, rather than a limitation of the analytical methodology itself.

Platform analysis

Sentiment analysis relies not only on the effectiveness of models but also on rich and diverse content that needs to be analyzed. At this stage of the platform lifecycle, the volume and distribution user-generated discussions do not yet allow for significant sentiment interpretations across all thematic areas.

Consequently, while research and code development for sentiment analysis have been conducted, platform analysis would require a greater volume of data to be meaningful in terms of extracting definitive indicators of public opinion. Methodological integrity is applied, while allowing the sentiment analysis infrastructure to evolve with the organic growth of platform interactions.

5.2 Gamification Incentives

ITHACA's gamification component aims to motivate users of the platform to participate as actively as possible and to maintain this motivation in the long term. Gamification principles (and actual implementations) are often criticised for promoting only extrinsic motivation, i.e., motivation to be rewarded with points, badges, etc. Gamification principles can indeed be very successful in promoting extrinsic motivation. However, intrinsic motivation, i.e., motivation derived from the activity itself (as in the case of ITHACA: interest in the topics, in participatory involvement, in exchange with others), is often neglected. In order to create the best possible conditions for extrinsic motivation to be converted into intrinsic motivation in the long term, we have started from a conceptual point of view with the Self-determination Theory[4]. For a description of its main principles, and how we used this theoretical foundation to make decisions about the specific design of the gamification incentives and underlying rules we would like to refer to D3.1 (Chapter 7).

As a summary, we decided to include the following for ITHACA users:

- A points and badge system based on user engagement levels:
  - Activity badge (e.g., login frequency, votes cast),
  - Competence badge (e.g., proposals submitted, comments written),
  - Relatedness badge (e.g., collaborative actions, helping others).
- Leaderboards (that are optional and anonymizable to preserve user agency and avoid negative social pressure).
- Clear “how to earn” explanations to ensure transparency and additionally foster intrinsic motivation

Since the submission of D3.1 (and D3.2), where the primary focus was still on theoretical elaboration and conceptualization, the primary focus has since shifted to actual implementation and evaluation (see also section 7.1 in the report, as well as D4.3 for details) activities.

In order to make it as clear as possible what adjustments were made between the conceptualization (D3.1) and the final implementation of the gamification incentive mechanisms (this report), the

[4] https://selfdeterminationtheory.org/theory/

following two tables are taken from D3.1, but clearly show how the initial concepts) have been adapted since then (modifications are colored in green, while missions that were omitted from the final version of the gamification component are colored in orange). Nevertheless, the theoretical foundation remained the same.

Table 1: List of activities (clustered into easy, medium, and difficult) that can be carried out (and thus, rewarded by badges) only a single time

Badges for the acquisition of single activities	
Bronze	*Create your 1st post (Competence) *Create your 1st comment (Competence) *Navigate/Engage through X component (X being argument visualization, simplification tool, TSD, evaluation tools, etc. (Activity) -> Mission "Engage with all Components!!!! - Engage with all components of the platform." (Activity).
Silver	*Update "About Me" question. (Relatedness) -> The "About Me" section was not implemented in the user's profile in the final version of the ITHACA platform. *Complete all Bronze level missions (Competence) -> This mission was modified as the "Relatedness" mission, "Explorer's Degree - Complete all single Acquisition Missions".
Gold	*Have your proposal passed on a civic topic (Relatedness) -> This mission was incorporated into the "Competence" mission, "Positive Proposal - Write a proposal & receive 20 positive evaluations.". *Have your argument be the most voted one in a civic topic (Relatedness) -> This mission was integrated in the mission "Proposal Master - Create 5 proposals & receive 50 total votes." (Relatedness).

Table 2: List of activities worthy of reward, their 'difficulty level', and the badges (or types of basic needs they should facilitate

Badges for the acquisition of repeated activities			
	Badge on Competence	Badge on Relatedness	Badge on Activity

Easy (1-10 Points); required to reach bronze status	*The user evaluates a comment of another user (e.g., by thumbs-up). *The user participates in a voting.	*The user associates him-or herself with other users (i.e. getting 'friends') *The user forwards a comment or proposal to others who might be interested in it.	*Number of logins *Amount of time spent on the platform. *Browse through the 5 most popular X (arguments, posts, etc.) *Browse through the 5 most recent X (arguments, posts, etc.)
Medium (11-25 Points), required to reach silver status	*The user writes a reply to a comment of another user. *The user writes a comment. *The user gets positive evaluations from other users for his/her comment. *The user who replies with a constructive comment to another users proposal receives an approval-rating (thumbs-up) by the the user who wrote the proposal (by this, the proposal-writer could be rewarded because he or she is 'allowed' to reward associated contributions / comments of other users that might even improve the initial proposal, and this would reward 'constructive' comments/ replies) -> This mission was changed into the mission "Comment on a proposal - Comment on a proposal & receive 5 positive evaluations." (Competence).	*The user answers a 'how-to' question of another (new) user. -> Since the "About Me" section was not integrated in the user's profile, the "how-to" question was also not implemented, hence this mission was omitted from the final version of the Gamification System.	*Make X posts with a positive feedback ratio (week duration) *Engage with X component (e.g. start a discussion/voting in argument visualization with positive feedback) -> Mission "Engage with all Components!!!! - Engage with all components of the platform." (Activity).

Difficult (26-50 Points), required to reach gold status	*The user writes a proposal (e.g. ideas for the city, such as how public transport could be improved). *The user does some research on city-related solutions and approaches from another (foreign) European city. For example: A user visits another European country / city. He or she sees a nice playground. He or she makes a photo and posts it, together with a description on why he or she thinks that this could be adapted to his/her hometown as well. -> This mission was implemented as the mission "Positive Proposal - Write a proposal & receive 20 positive evaluations." (Competence).	*The user initiates a (successful) voting; whereas successful means that a lot of other users participate in the voting. -> This mission was modified as follows: "Proposal Master - Create 5 proposals & receive 50 total votes." (Relatedness). *The user writes a proposal (see also cell at the left) together with others - i.e. collaboratively and cooperatively. Up to a certain group size, e.g. 5 users, the points should be equally distributed and as high as if the user would do it alone. Above that, the 'individual reward' (i.e., the number of points a single user of the group gets) should be reduced accordingly. -> The final version of the Argument Visualization Tool did not support collaborative editing of proposals, thus this mission was omitted.	

Since the Argument Visualization Tool was developed as the main forum, where users discuss public issues, the term "post/s" refers to the proposal/s created to disseminate opinions and democratic solutions regarding these issues. Moreover, all missions outlined in Tables 1 & 2 were implemented as missions that can be completed only once, to prevent users from exhibiting highly competitive or "addictive" behaviors (e.g., "chasing" rewards, obsessing over badge collection, competing against other users on how many or how quickly they complete a mission, etc.). More details regarding the available missions on the Gamification Engine are provided in Tables 1, 2 & 3 in Chapter 5.2 of D3.4.

As briefly mentioned above, we would like to emphasize once again that UPAT conducted its own evaluation study on the gamification component. An independent evaluation was deemed appropriate because including the gamification component evaluation with the

other phase 2 evaluation aspects was considered too extensive and time-consuming for the potential end users. The results and conclusions are described in section 7.1 in this report, as well as in more detail and including methodological considerations in D4.3.

5.3 AI-powered functionalities

The ITHACA platform integrates a set of AI-powered functionalities designed to support inclusive, informed, and constructive civic participation. These services operate as modular components, exposed through secure APIs and seamlessly embedded into user workflows. In line with the principles defined in earlier deliverables, particularly D3.3, all AI functionalities are designed to be **assistive rather than decisive**, maintaining human oversight, transparency, and user control.

5.3.1 Summarization

The **Summarisation tool** assists users in navigating large volumes of user-generated content by automatically generating concise summaries of discussions, proposals, and argument threads. Its primary objective is to reduce cognitive overload and enable citizens, moderators, and municipal stakeholders to quickly grasp the main points of ongoing civic debates.

The summarisation functionality is applied at multiple levels of the platform:

- **Topic-level summarisation**, where the system produces an aggregated overview of the discussion surrounding a civic issue.
- **Proposal- and argument-level summarisation**, allowing users to obtain short, readable digests of long or complex contributions.

The tool is implemented using Natural Language Processing (NLP) techniques capable of handling multilingual input, reflecting the linguistic diversity of the pilot sites. Summaries are generated on demand and are clearly marked as AI-generated to ensure transparency. Users remain free to consult the original content at any time, reinforcing the supportive role of the AI.

In accordance with the trust and accountability principles described in D3.3, summarisation does not alter or suppress content. Instead, it provides an additional interpretative layer aimed at improving accessibility, especially for users with limited time, cognitive constraints, or lower familiarity with formal civic language.

5.3.2 Translation

The **Translation tool** enables multilingual participation by allowing users to translate platform content across the supported languages of the pilots (English, Romanian, and Slovak). This functionality is essential for fostering inclusiveness and ensuring equal access to civic discourse regardless of users' linguistic background.

Translation is available for:

- Topics, proposals, and arguments posted by other users.
- AI-generated summaries.
- User-authored content prior to submission, enabling users to verify translations before publishing.

The translation service is based on state-of-the-art neural machine translation models and is tightly integrated into the user interface through contextual action buttons. Translated content is displayed alongside or in place of the original text, with clear visual indicators that the output is machine-generated.

As highlighted in D3.3, the translation tool is designed to support understanding rather than replace original expressions. Users are encouraged to engage with content in their preferred language while preserving the original meaning and intent of contributors. This approach proved particularly valuable during pilot validation, where cross-language interaction between participants was a key requirement.

5.3.3 Toxicity detection / moderation

The **Toxicity Detection and Moderation tool** plays a central role in ensuring respectful, safe, and constructive civic dialogue on the ITHACA platform. The tool employs AI-based text classification models to identify potentially inappropriate, offensive, or harmful content submitted by users.

Toxicity detection is applied at critical interaction points, including:

- Submission of proposals.
- Posting of arguments and comments.
- Free-text interactions with AI services.

The moderation process follows a human-in-the-loop (HITL) approach, as defined in D3.3:

- Content identified as non-toxic is published immediately.
- Content flagged as potentially toxic is either blocked at submission time (for severe cases) or routed to moderators for review.
- Moderators retain full authority to approve, edit, or reject flagged content.

Thresholds and confidence levels are configurable, allowing adaptation to contextual and cultural sensitivities identified during the pilot phase. Importantly, users receive feedback when content is flagged, supporting transparency and enabling them to revise submissions rather than being silently censored.

This hybrid moderation model balances automation efficiency with democratic accountability, ensuring that AI supports—rather than replaces—human judgment in civic discourse.

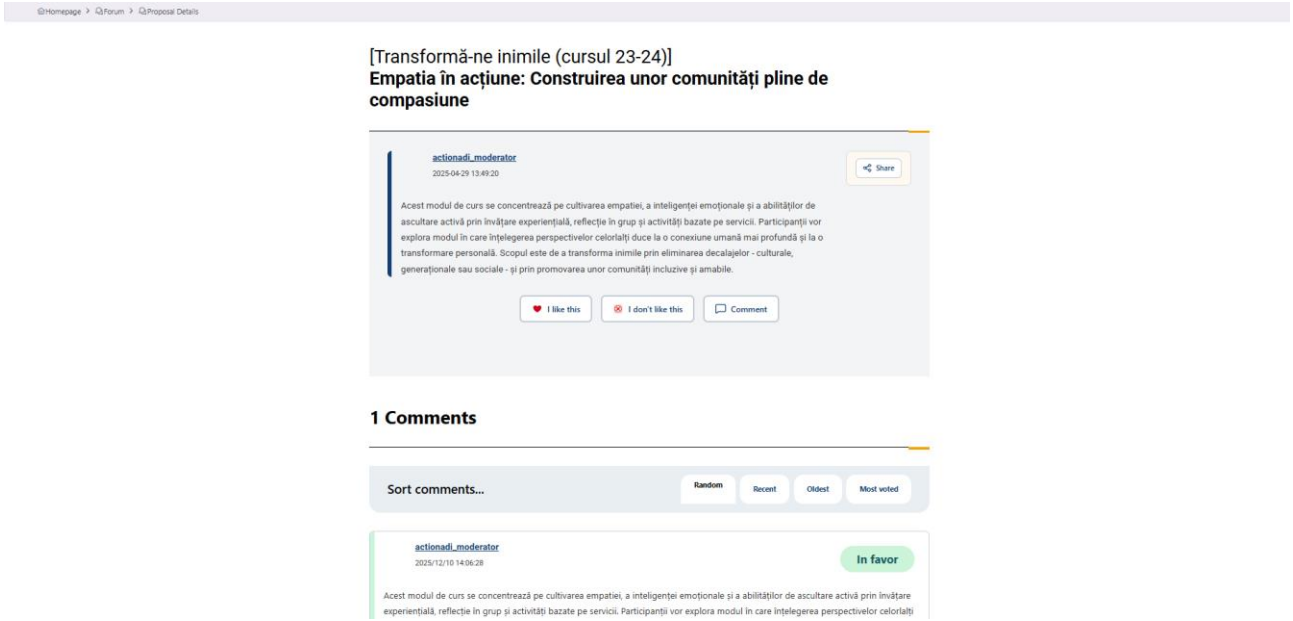
5.4 Argument Visualization (UPAT, SIMAVI)

The Argument Visualization component is designed to present user suggestions and arguments regarding civic issues in a structured and visually informative manner. This functionality goes beyond simple commenting by organizing contributions into proposals and associated arguments, supported by visual analytics to gauge community sentiment.

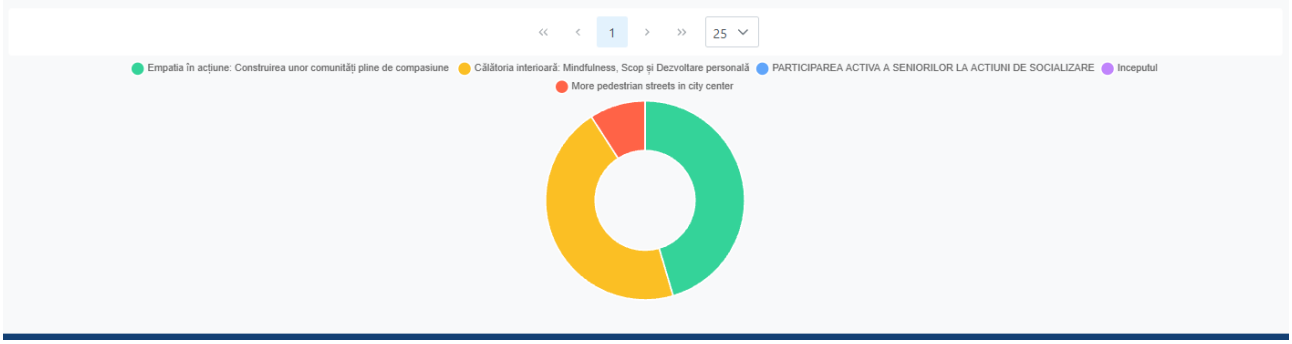
5.4.1 Technical Implementation

The Argument Visualization tool is implemented as a distinct module within the ITHACA platform's frontend (Angular) and backend (Spring Boot) architecture.

- **Data Model:** The core entity is the **Proposal**, which is linked to a specific **Topic** (Issue). Each proposal contains a title, description, author information, and a list of associated **Arguments**. Crucially, the data model supports granular voting: users can vote on the proposal itself as well as on individual arguments.
    - **Entities:** `Proposal`, `Argument`, `Topic`.
    - **Voting Logic:** User votes are tracked per entity (proposal or argument). The system calculates a net vote score (`upvotes - downvotes`) to determine the ranking and visual representation of each item.
- **Frontend Integration:**
    - The user interface provides a dedicated **"Arguments" tab** or section within the forum.



- Proposals are displayed as distinct cards, allowing users to drill down into the specific arguments supporting or opposing each idea.
- **Interactive Charts:** A dynamic **Pie Chart** (implemented using `primeng/chart`) visualizes the distribution of votes across different proposals for a given topic. This provides an immediate, at-a-glance understanding of which ideas are gaining the most traction within the community. The chart updates in real-time as users cast their votes.



- **Backend Services:**
  - **Proposal Service:** Handles the creation, retrieval, and deletion of proposals. It exposes REST API endpoints (e.g., `GET /proposals/issue/{issueId}`) that the frontend consumes to populate the visualization.
  - **Voting Service:** Manages the logic for casting and tallying votes, ensuring data integrity and preventing duplicate votes from the same user.

5.4.2 Visualization Logic

The visualization logic aims to reduce bias and promote objective evaluation of ideas.

- **Randomized Display:** To mitigate the "bandwagon effect" where users simply vote for the top item, proposals can be displayed in a randomized order initially. Sorting options (e.g., "Most Voted", "Recent") are available for users who wish to explore the data more systematically.
- **Net Vote Calculation:** The system calculates a `netVote` value for each proposal. This value is used to generate the data for the pie chart, ensuring that the visual representation accurately reflects the community's consensus (or lack thereof).
- **Color Coding:** Proposals in the visualization are assigned distinct colors to make the chart easily readable. A predefined palette of pleasant colors is used, with a random color generator as a fallback for topics with a large number of proposals.

This technical approach ensures that the Argument Visualization tool is not just a static display but an interactive, data-driven component that enhances the deliberative quality of the ITHACA platform.

5.5 Trust & Security Infrastructure (SIMAVI)

The Trust and Security Infrastructure forms the foundational layer ensuring the **confidentiality, integrity, and availability** of the ITHACA platform and its sensitive civic participation data. Building upon the architectural decisions detailed in **D3.1 (System Architecture)**, SIMAVI implemented a robust security framework focused on Authentication, Authorization, and Data Protection.

5.5.1 Authentication and Authorization

The security architecture employs a centralized approach to Identity and Access Management (IAM), relying on **Keycloak** as the authoritative provider.

- **Single Sign-On (SSO):** Keycloak manages all user identity lifecycles and provides a cohesive Single Sign-On experience across all platform microservices and interfaces.
- **Role-Based Access Control (RBAC):** The platform enforces a strict RBAC model. Keycloak is configured to assign specific roles (e.g., `Citizen`, `Moderator`, `Admin`) which are used by the application to grant or deny access to resources, data, and critical functionalities (such as moderation dashboards or AI model controls).
- **OpenID Connect (OIDC) / OAuth 2.0:** All APIs are secured using the OIDC/OAuth 2.0 protocol. Upon successful login, Keycloak issues a **JSON Web Token (JWT)**. This token is attached to every API request and is validated by the receiving microservice for both authentication (proof of identity) and authorization (proof of required role/permission).

5.5.2 Data Security and Confidentiality

Data handling adheres to high standards of confidentiality to comply with the project's ethical guidelines and relevant regulations (e.g., GDPR).

- **Transport Layer Security (TLS):** All data exchange, including frontend-to-backend communication and internal service-to-service calls, is encrypted using **HTTPS/TLS 1.2+**. This mechanism prevents man-in-the-middle attacks and ensures data integrity during transmission.
- **Data Minimization and PIMS:** The security model is aligned with the principles of the **Personal Information Management System (PIMS)** established in coordination with other partners. This ensures that PII is collected only when strictly necessary, stored securely, and anonymized or pseudonymized where feasible to protect user privacy while supporting analytic requirements.
- **Secure Storage:** Databases and other persistent storage components are deployed with encryption-at-rest enabled to protect stored data. Access is strictly controlled via isolated network segments and dedicated, least-privilege service accounts.

5.5.3 Infrastructure Security

SIMAVI secured the operational environment by focusing on container hardening and disciplined secrets management.

- **Container Security:** Microservices are containerized using **Docker**. Images are built based on minimal base operating systems (e.g., Ubuntu Linux, as referenced in D3.1) to minimize potential vulnerabilities and reduce the overall attack surface of the deployment.
- **Secrets Management:** Critical credentials and sensitive configuration parameters (database passwords, API keys, cryptographic secrets) are not stored in environment variables or code. Instead, a dedicated **Secrets Management** utility or pattern is used to securely inject these secrets into the application containers at runtime.
- **Network Segmentation (Deployment Environment):** The microservices operate within a managed, tightly controlled network environment. Network access is restricted by policy, ensuring only explicitly required ports are open for external communication and only necessary inter-service communication paths are permitted. This layered defense helps isolate services and prevent unauthorized traversal of the network.

5.6 Personal Information Management System

5.6.1 Overview and Purpose

The Personal Information Management System (PIMS) is intended to act as a core module of the platform, helping to ensure data protection guidelines and build user trust via transparency, control, and accountability. The PIMS is not designed as a standalone system, but rather as a module integrated with the platform's architecture, interacting with the platform's AI components and content management processes. Its primary objective is to help users manage and control how their personal data is processed and shared throughout the platform.

5.6.2 Integration

The Personal Information Management System follows the existing process of user authentication and identity management being handled by an external system, relying on Keycloak as the authentication provider. As a result, the PIMS does not manage user identities directly and does not store authentication credentials, in order to minimize data collection points and retain as little data as possible. Instead, it trusts the externally-issued authentication tokens and user identifiers and uses them as references during handling data protection actions.

The PIMS acts as a policy-enforcement layer, receiving authenticated requests from users, verifying their permissions based on trusted identity claims, and turning user privacy directives into actions across the platform's internal services. This design allows for modularity and enhancement or refinement of actions.

5.6.3 Core Functionalities and User Actions

The PIMS provides users with the ability to define how their personal data may be used and by which components of the platform. A central function of the PIMS is consent management. Users can express consent for specific categories of data processing, such as the analysis of their comments for sentiment analysis.

An aspect of this consent withdrawal. When a user withdraws consent, this is recorded by the PIMS and evaluated in the internal levels of the platform. This may involve disabling future AI analysis of user content, excluding the user's data from new analytical results.

Another core capability is user-initiated data actions, such as requesting the deletion of their own content. In such cases, the PIMS verifies the user's authority over the data, identifies all relevant data objects (e.g. comments, discussion posts), and calls the appropriate platform services to perform deletion or anonymisation.

5.6.4 Interaction with AI Components

The PIMS is designed to interact with the platform's AI tools, acting as a controller that determines whether specific AI processes are permitted to operate on user data. Relevant AI services are expected to verify through PIMS whether a user has granted consent for data processing. This ensures that data protection principles are enforced consistently across automated pipelines.

However, there are limitations by design: for example, in order to provide a safer deliberation space, users cannot opt out of their submitted content being processed by the toxicity checking tool, due to concerns of safety being compromised by potentially malicious users that may opt out from this check, intending to post inappropriate content. Therefore, user control is limited to AI functionalities that do not affect the platform's scope of operations or functional design. A safety measure that has

to be noted is that all AI components inside this platform are self-hosted by the platform and only process data internally, so that no third-party data exchange occurs.

5.6.5 Flow of actions

From a user perspective, interaction with the Personal Information Management System happens as follows: After authenticating through the external identity provider, the user accesses the platform and navigates to a dedicated area for data and privacy controls. Within this space, the user can review existing consents, modify permissions, or initiate actions such as data deletion requests.

Each user action is validated against their authenticated identity and translated into a request handled by the PIMS. The PIMS then determines which platform components need to be activated in order to complete the action. Throughout this process, the system maintains an internal record of actions taken, enabling future audits.

Overall, the Personal Information Management System constitutes a foundational element for trustworthy civic technology, reinforcing user agency and legal compliance within the ITHACA platform.

5.7 Continuous Integration & Deployment Pipelines

The deployment strategy for the ITHACA platform is built upon a **Container-Based Artifact Delivery** model. Jointly managed by **KT** and **SIMAVI**, this approach prioritizes system stability, ease of maintenance, and rapid debugging within the single operational environment, ensuring a consistent transition from development to production.

5.7.1 Build Process and Artifact Generation

Rather than a monolithic pipeline, the platform utilizes a modular build strategy where application artifacts are generated and validated before containerization. This ensures that only successfully compiled code is packaged for deployment.

- **Backend Services:** The Java/Spring Boot microservices (Backend, Communication, Orchestrator and Audit layers) are compiled into executable JAR files. These are packaged into Docker images based on the **bellsoft/liberica-openjdk-debian:17** runtime, ensuring a secure and optimized Java 17 environment.
- **Frontend Application:** The Angular-based user interface is pre-built into a distribution (dist) package. To minimize resource footprint, the deployment image uses a lightweight **node:16-alpine** base, serving the static content via the high-performance **serve** utility on port 4200.
- **Debug Readiness:** To support the rigorous testing phases defined in WP4, the backend container images are configured with **remote debugging enabled** (exposing ports **5005** and **5006**). This allows developers to perform deep-dive diagnostics and "hotfix" analysis directly on the running environment during Phase 0 (Internal Verification).

5.7.2 Continuous Deployment (CD) and Evaluation Phases

The deployment model operates on a single, highly controlled operational environment hosted on Virtual Machines (VMs). Given this architectural choice, the CD process focuses on reliable, atomic updates to the live system, which then supports a sequential testing strategy involving widening circles of users.

- **Deployment Mechanism:** Deployment is managed using **Infrastructure as Code (IaC)** principles, primarily through customized **Docker Compose** configurations and dedicated shell scripting tailored for the host VMs. This approach allows for repeatable deployment steps and atomic updates of specific services (e.g., updating the backend logic without disrupting the database) without requiring a complex orchestrator like Kubernetes.
- **Phased Evaluation Strategy:** To manage risk within this single environment, the platform updates support a structured evaluation timeline defined in **D4.1**. Rather than separate deployment environments, the live platform undergoes progressive validation by distinct user cohorts:
    - **Phase 0 (Internal and partner's evaluation):** Immediately following a deployment, access is prioritized for the development teams (**KT, SIMAVI**) to perform technical checks and validate system stability using the exposed remote debugging tools.
    - **Phase 1 (User testing and results):** Once technically verified, the platform functionality is validated by consortium partners and Quality Assurance personnel. This phase ensures systemic stability and adherence to specifications before opening the system to external users (D4.2).
    - **Phase 2 (User testing and results):** The stable, validated code is then exposed to extended user groups from the pilot cities (**Brașov, Martin**) for **User Acceptance Testing (UAT)** and feedback gathering, as detailed in the evaluation reports (D4.3)

6. Platform Integration and Operation

6.1 User workflow (from login to participation)

6.1.1 Homepage

The SIMAVI platform's homepage provides a comprehensive entry point to all interactive features and is designed to be fully accessible and browsable **without requiring a user to be logged in**. Authentication is only required for interactive functions and accessing the personalized profile data.

Here is a walkthrough of the main sections observed on the homepage:

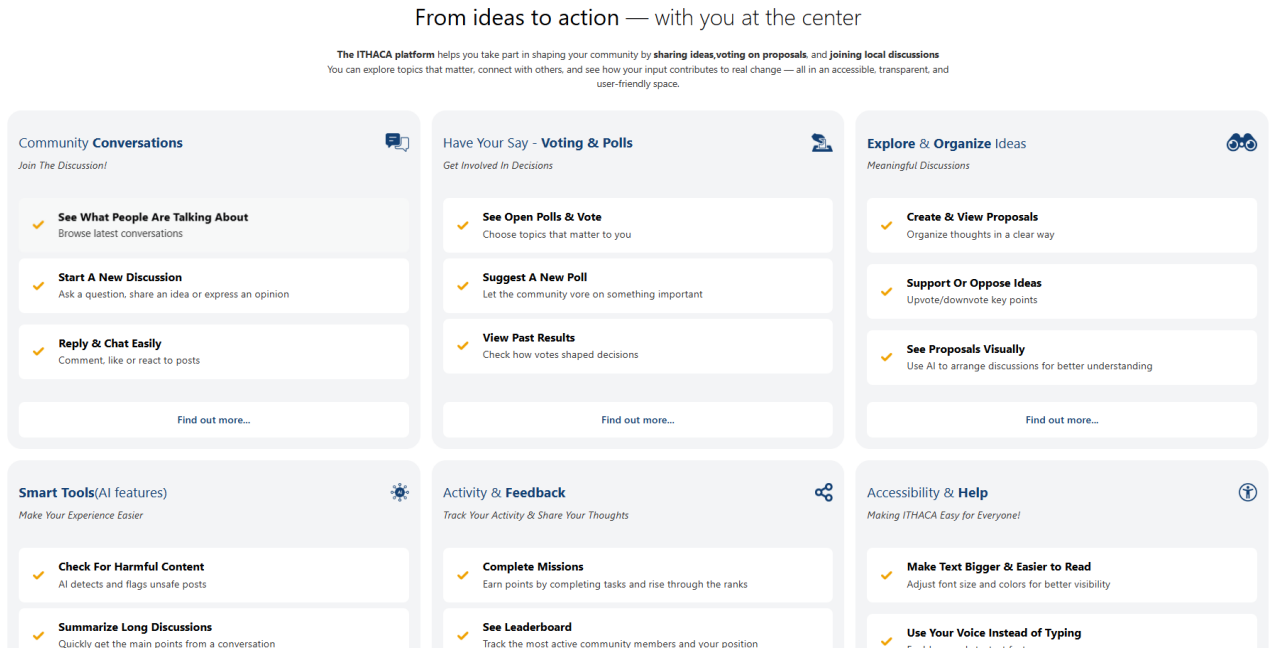
1. Main Banner and Welcome

- **Welcome Message:** The page opens with a large banner displaying the project name, **ITHACA**, and the tagline: "Inclusive Technology for Human-Centric AI Civic Participation."
- **Participation Options:** Directly below the banner, four prominent calls-to-action guide the user toward interactive modules:
    - **Start a Discussion -** first action card redirects to the topics of discussions, allows the users to explore and share ideas with the community
    - **Create a Poll or Vote -** second action box redirects to polls section, allows the user to contribute to collective decision-making
    - **Test AI Tools -** 3rd action box, allows the user to get familiarised with the ai tools that are used.
    - **Chat with Our Assistant -** 4th action box, allows the user to interact directly with the AI support system

2. Join Conversation (Featured Topics)

This section highlights currently active discussions and serves as a gateway to the community forums.

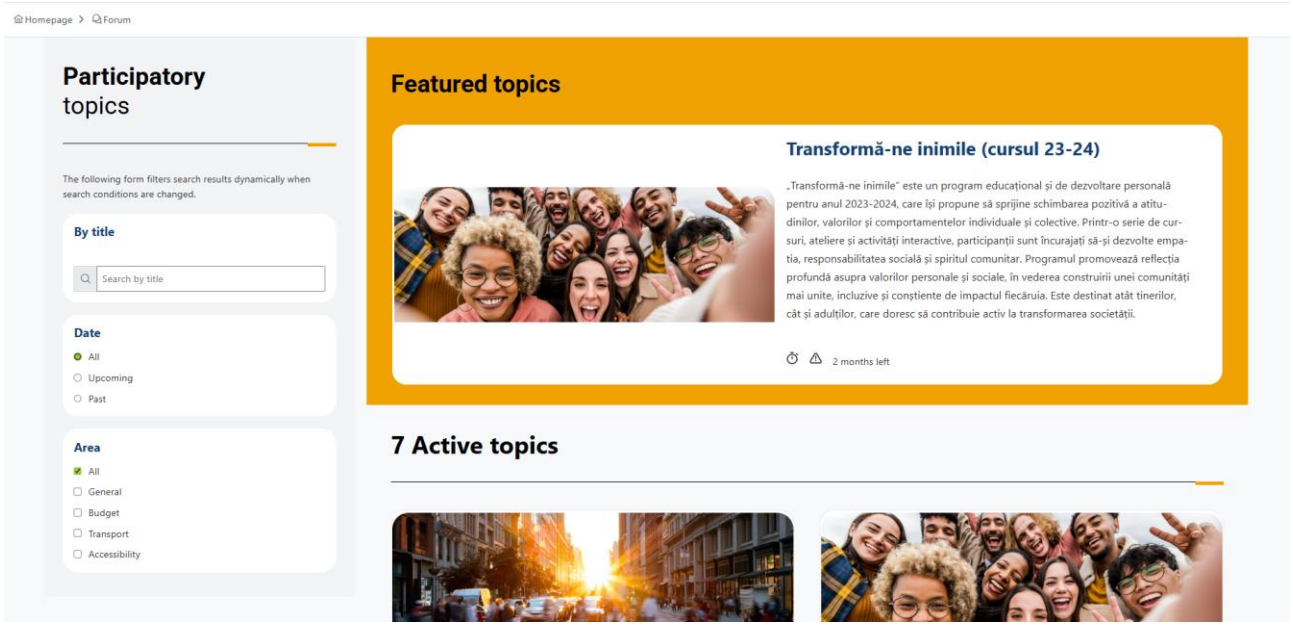
- It features a visual grid of active topics (e.g., "Transformă-ne inimile," "Parcuri de Locuințe") inviting the user to explore ongoing community conversations.



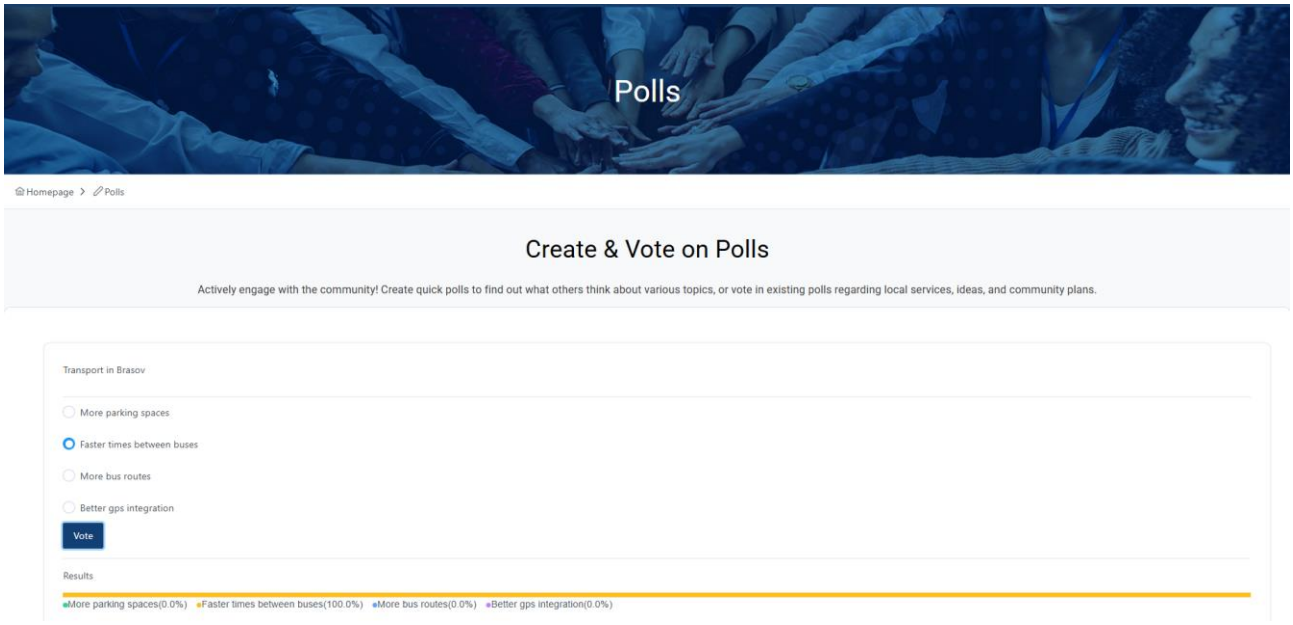
3. From Ideas to Action — With You at the Center

This central grid organizes the platform's features into six core functional pillars:

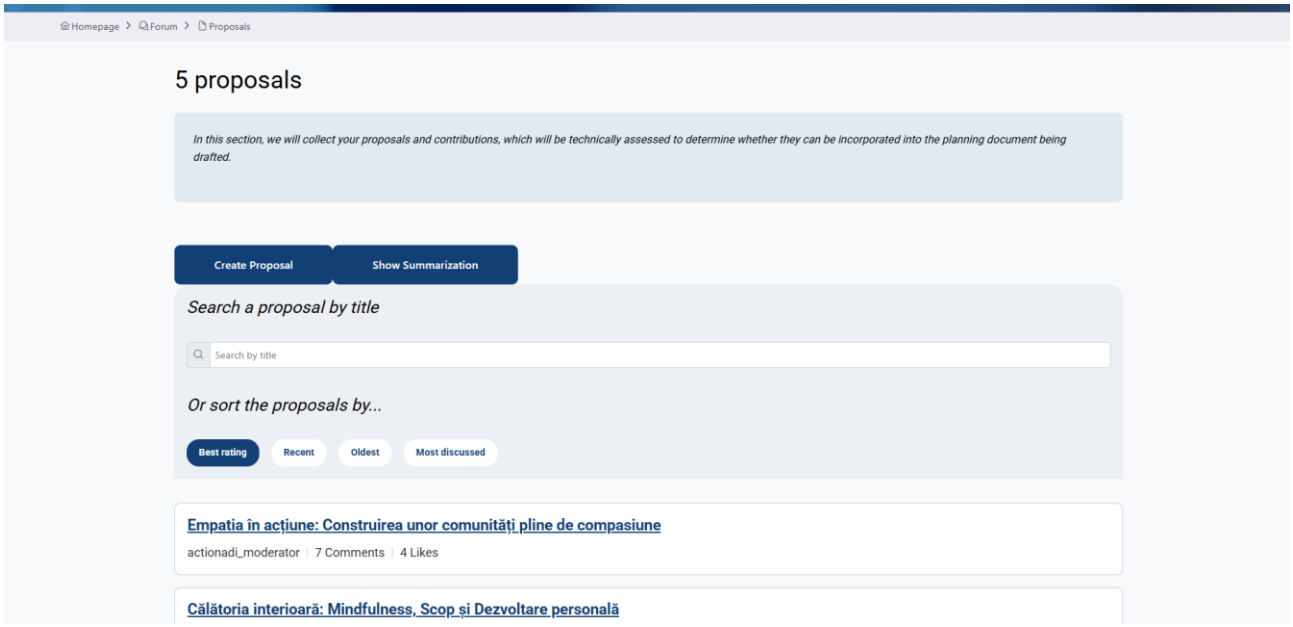
Community conversations | See what people are talking about * Start a new discussion * Reply and chat easily - You can join public conversations on topics that matter to your community. Simply click 'Start a discussion' button to read what others are saying or to share your own thoughts. You don't need to be an expert - your experience matters. Use plain language or even voice input to make yourself heard. (forum-section)



Have Your Say - Voting & Polls | * See Open Polls & Votes * Suggest a New Poll * View Past Polls - You can vote in local polls or surveys to help shape decisions. Just choose the option that fits your opinion. Each vote counts and helps your city know what matters to you. The interface is simple, with large buttons, high contrast, and support for screen readers. (polls)

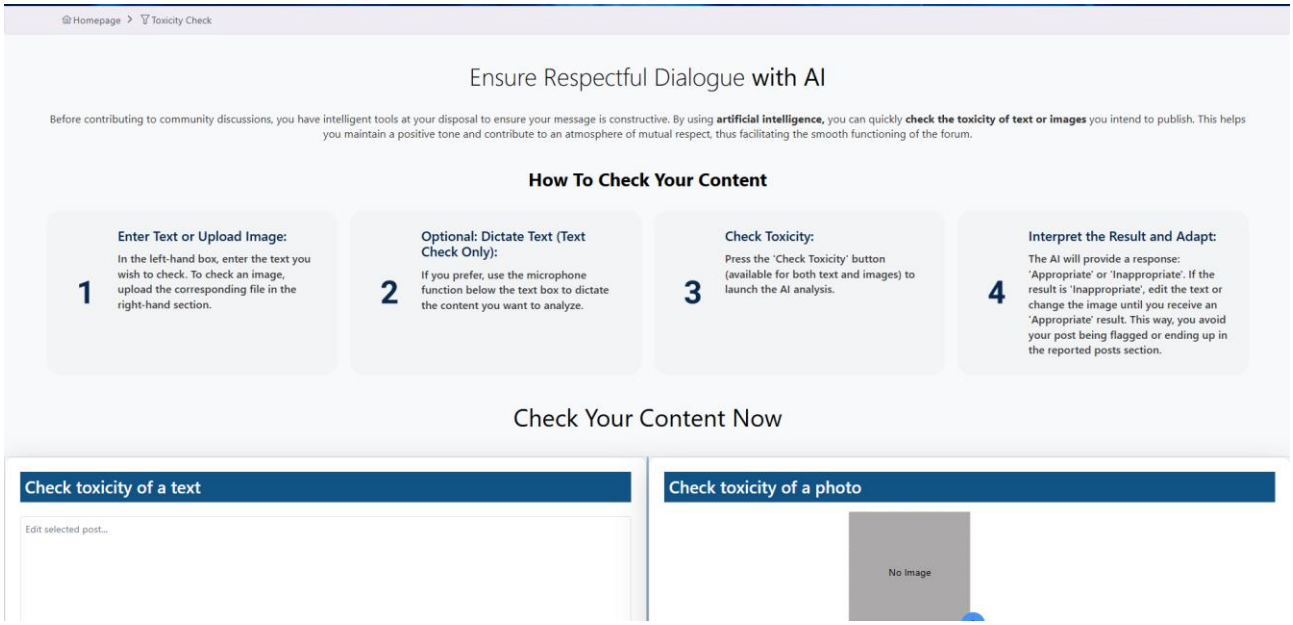


Explore & Organize Ideas | * Create & View Proposals * Support Or Oppose Ideas * See Proposals Visually - Browse ideas from your community and group them by topics like transport, health, or public spaces. (proposals)

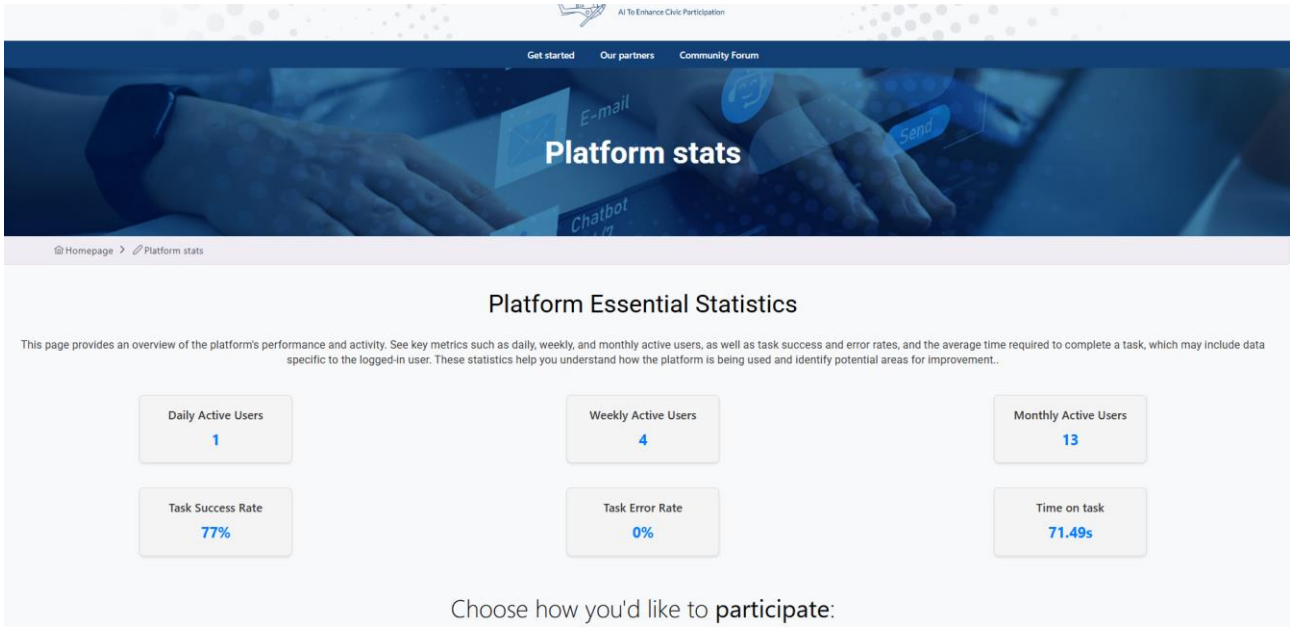


Accessibility & Help | * Make Text Bigger & Easier to Read * Use Your Voice Instead of Typing * Keyboard Accessibility * Get Help Anytime - ITHACA is designed to work for everyone. You can change text size, use voice commands, and navigate using a keyboard. There are also screen reader and translation options. If you ever need help, there's a friendly assistant and a help center.

Smart Tools (AI features): Highlights the AI/ML functionalities, such as checks for harmful content, instantaneous language translation, and automated text summarization.(ai-tools)



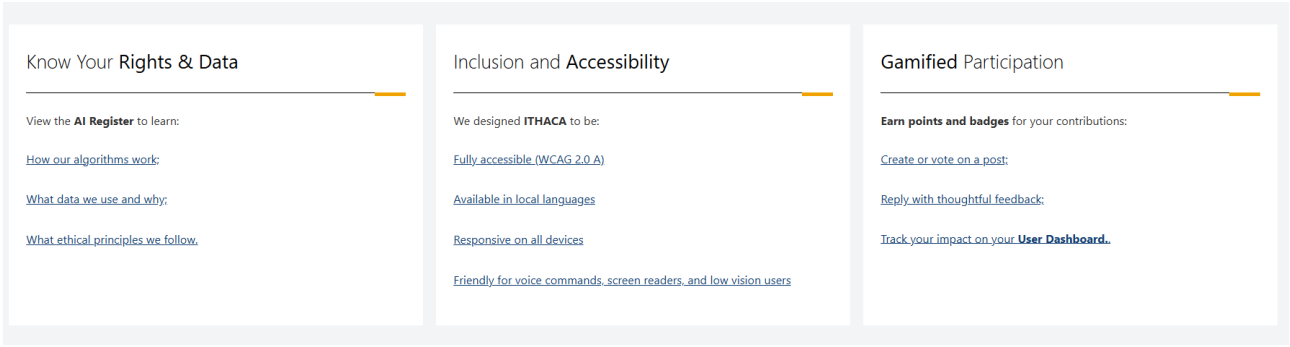
Activity & Feedback: Details how the platform tracks user interaction, including contribution levels, engagement statistics, and participation trends.[(stats)]



4. Auxiliary Features and Information

The bottom section of the homepage provides links to legal and technical assurances:

- **Know Your Rights & Data:** Links to information regarding user rights, data privacy, and intellectual property.
- **Inclusion and Accessibility:** Emphasizes the platform's commitment to accessibility, including support for vulnerable users.
- **Gamified Participation:** Describes the system for recognizing and rewarding user engagement through points and badges.



Authentication Requirements

Users are granted full viewing access to all content, including topics, proposals, visualizations, and general information, **without authentication**.

Interaction with the platform's advanced features, such as:

- Posting a new topic
- Posting a new proposal
- Voting on proposals or arguments
- Accessing the personalized **Profile Section** (where private profile data, like custom location and accessibility settings, is managed)

These actions require the user to be **logged in**.

6.1.2 Login and Register

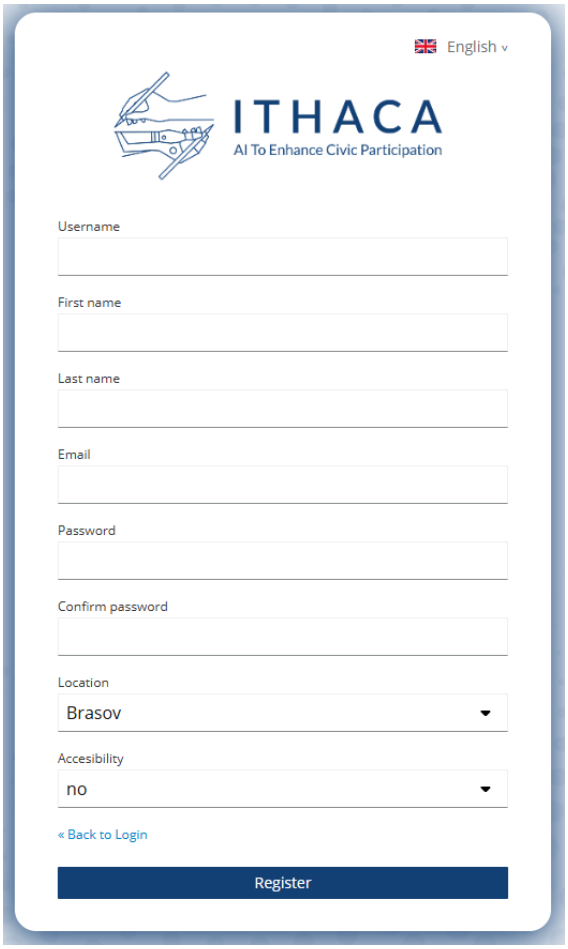
The platform uses **Keycloak** for authentication, featuring a customized interface for localized settings and accessibility.

1. Login Process

The login page provides access via:

- **Standard Credentials:** Username/email and Password.
- **Social Sign-In:** Options for Google and Facebook.

2. Registration and Profile Initialization



The registration screen requires users to create an account and define critical settings for the ITHACA platform:

- **Location:** Users must select their location (e.g., Brasov) to receive tailored content.
- **Accessibility:** Users specify their accessibility needs (e.g., "Yes" or "No"). This informs UI customization.
- **Minority Status:** An option is available for users to declare if they belong to a minority group.

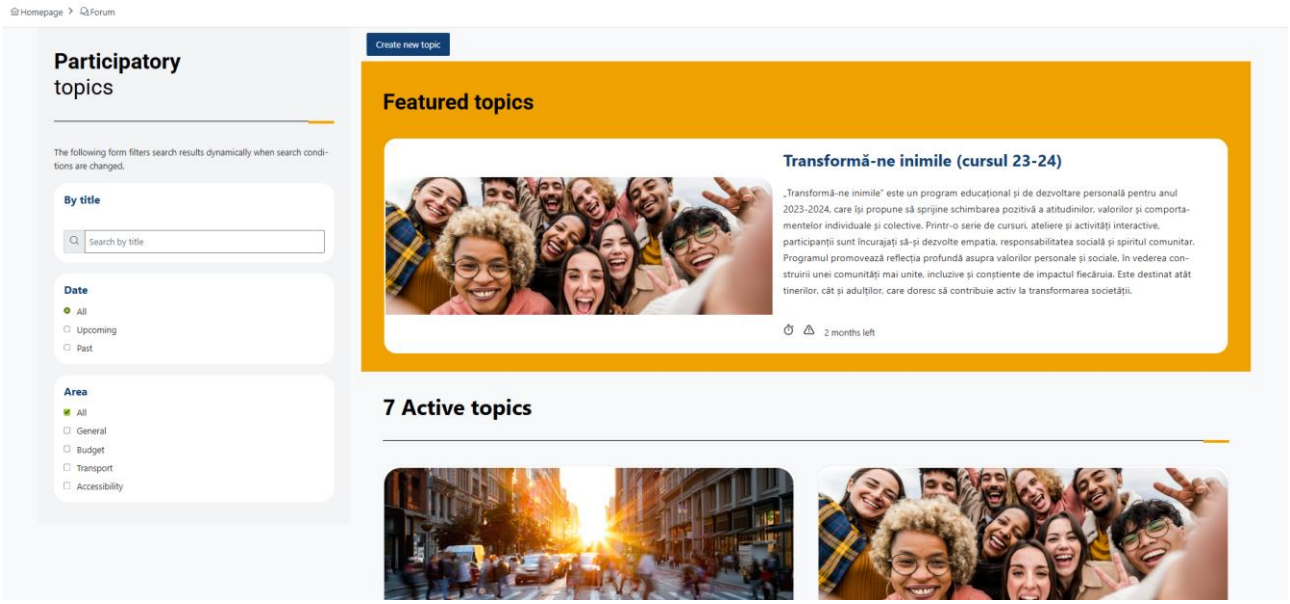
The process concludes with the user clicking **"Register"** ("Înregistrare") to finalize the personalized and inclusive profile setup.

6.1.3 Forum Access, Browsing, and Proposal Submission

This section covers browsing topics, evaluating proposals, and submitting new ideas.

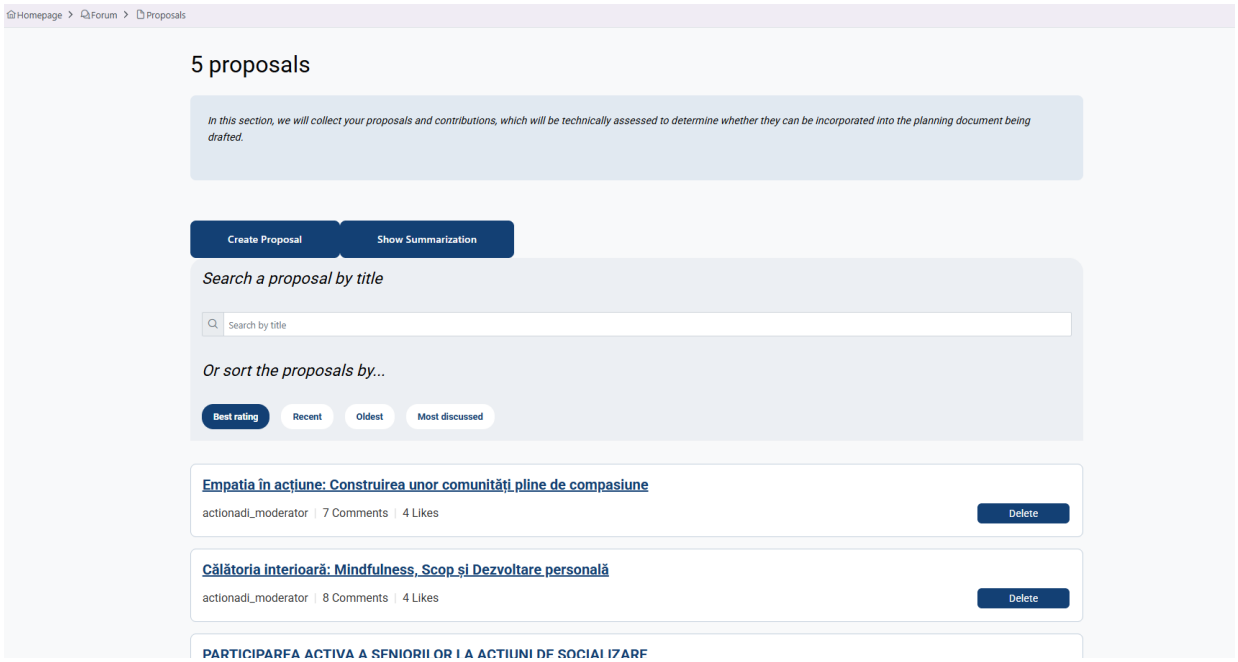
1. Forum Browsing and Topics

- **Access:** The forum is accessible via the "Community Forum" link on the main header.
- **Topic View:** Users view **"Participatory topics"** and can filter them by Title, Date (All, Upcoming, Past), and Area (General, Budget, Transport, Accessibility).
- **Moderator Action:** The **"Create new topic"** button is visible only to authenticated moderators.



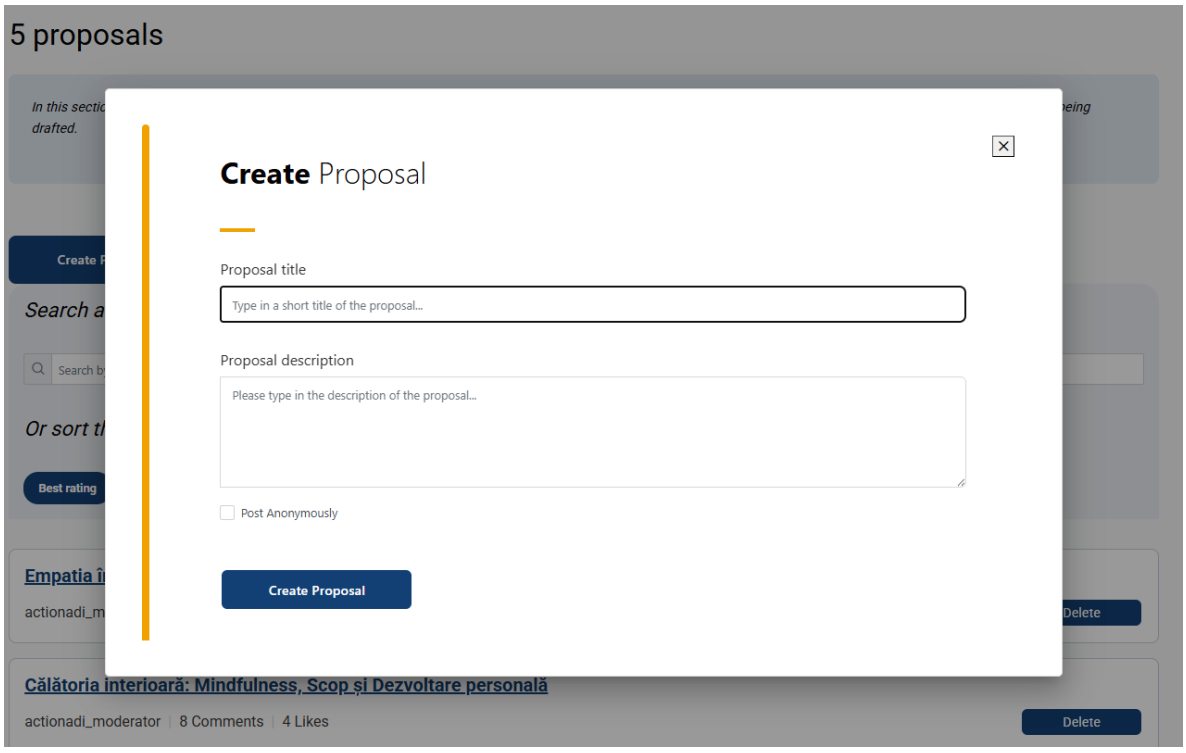
2. Proposals List and Evaluation

- **Access:** Clicking a specific topic leads to the Proposals page.
- **Filtering and Sorting:** Users can **Search** proposals by title and **Sort** them by criteria including "Best rating," "Recent," "Oldest," and "Most discussed".
- **Proposal Details:** Each proposal card displays the title, author (or 'Anonymous'), comments count, and total votes/likes. Moderators can see a **"Delete"** button.
- **Visualization:** At the bottom of the list, a **doughnut chart** visualizes the distribution of total votes, highlighting the percentage of support for each proposal.



3. Creating a Proposal

- **Access:** Authenticated users click the **"Create Proposal"** button to open a dialog.
- **Submission Form:** The form requires a **"Proposal title"** and a detailed **"Proposal description"**.
- **Anonymity:** Users have the option to check a box to **"Post Anonymously"**.
- **Toxicity Check:** The system checks the text for toxicity before submission to maintain a civil environment

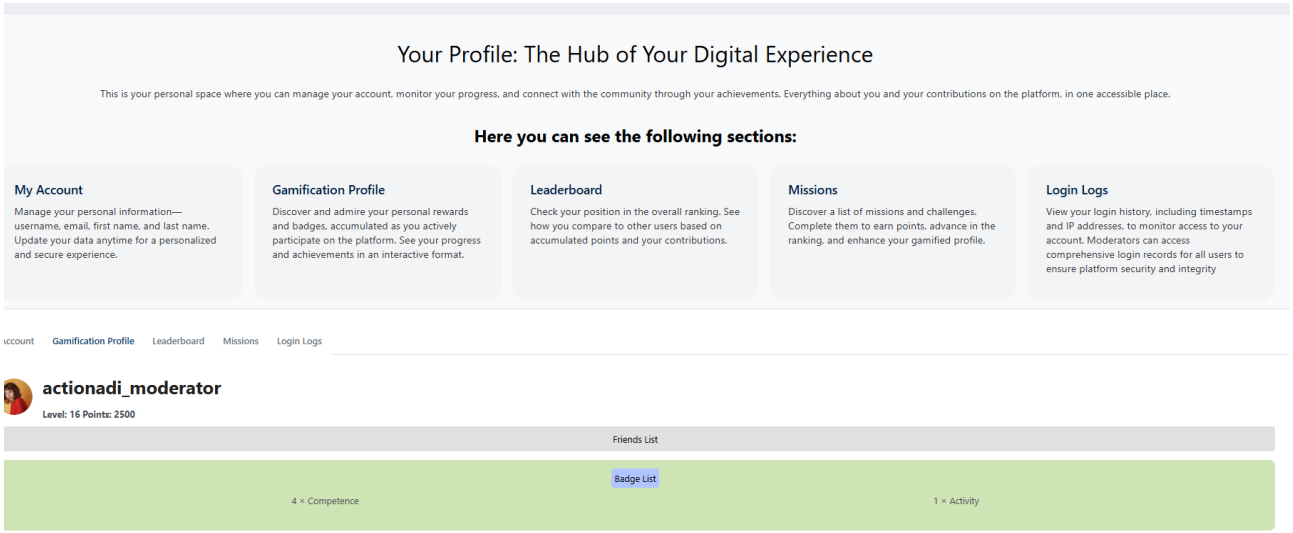


6.1.4 Gamification and Leaderboards

The platform uses a gamification system to track user activity, award points, and display community ranks to encourage continuous participation.

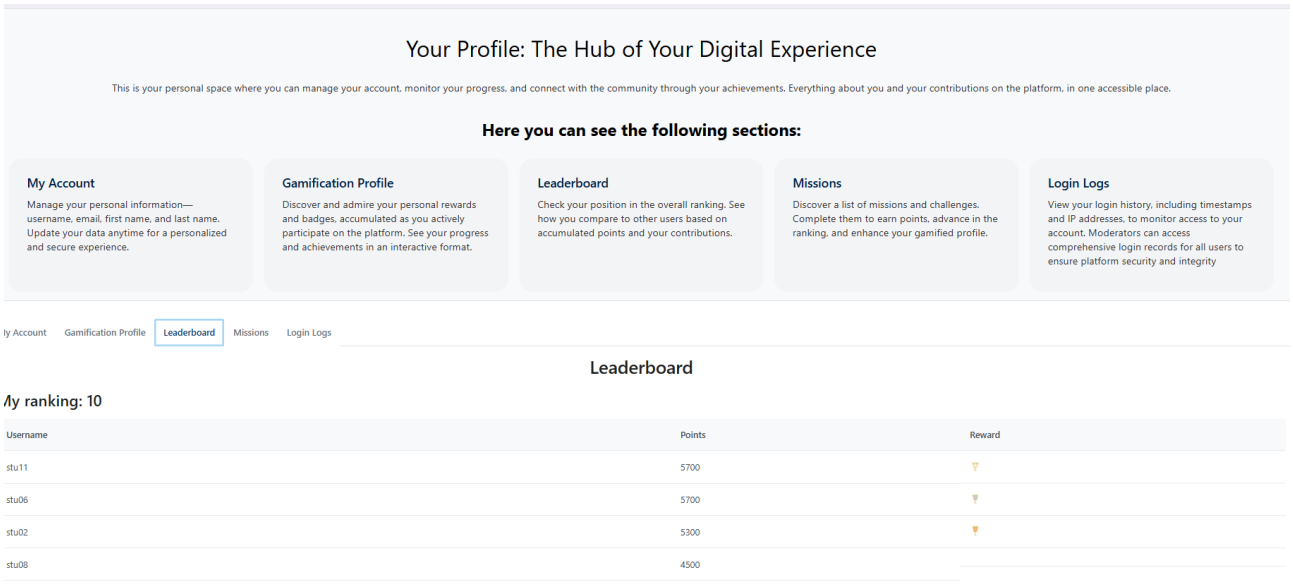
1. Gamification Mechanics

- **Experience Points (XP):** Users earn XP by completing various actions on the platform, such as logging in, posting proposals, writing arguments, commenting, and voting .
- **User Level:** XP accumulates to raise the user's level, which is displayed on their profile.
- **Badges:** Specific achievements are rewarded with badges , recognizing expertise or high activity in certain areas.



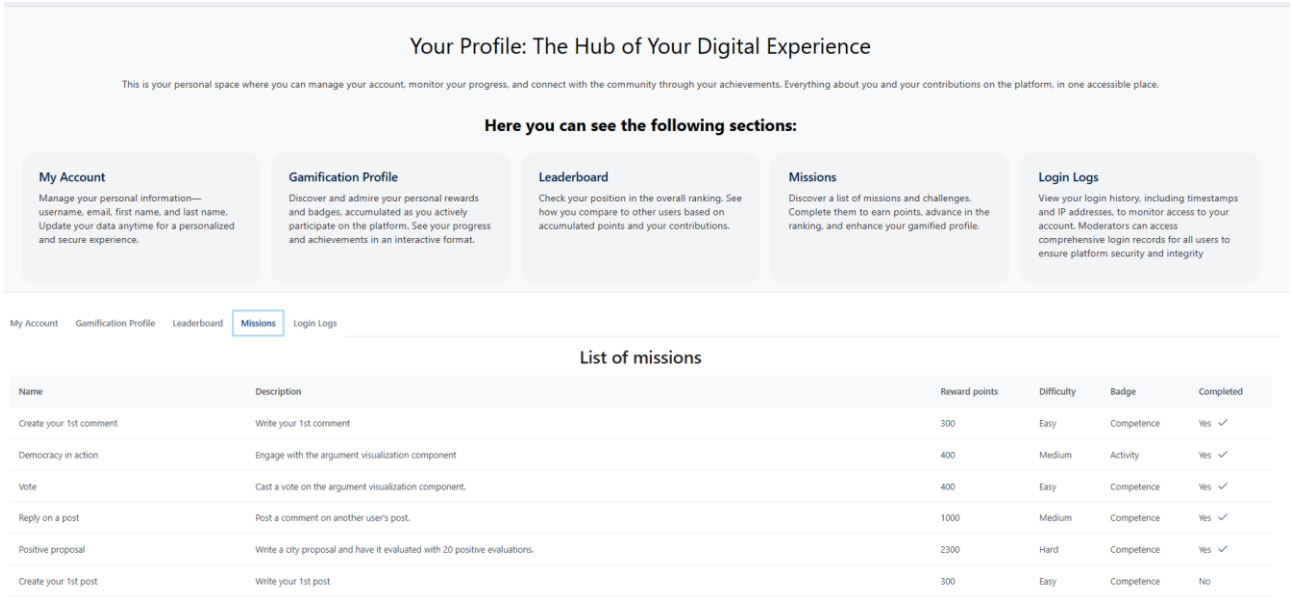
2. Leaderboards

- **Access:** Users can access the leaderboards from their profile .
- **Ranking:** The leaderboard displays users ranked by their total accumulated **Points** (XP).
- **Location Filter:** The ranking is filtered by the user's registered **Location** (e.g., Brasov or Martin), ensuring local relevance . The rank shows the user's position among others in the same city.
- **Top 10:** The primary view shows the top ten users by points in the selected location.
- **Personal Rank:** A user can query their specific rank within their location, even if they are not in the top ten.



3. Activity and Mission Data

Activity Tracking: The system tracks various participation metrics ("Mission Data"), including the number of arguments written, votes cast, time spent on the platform, and login frequency. This data forms the basis for awarding points and maintaining gamification.

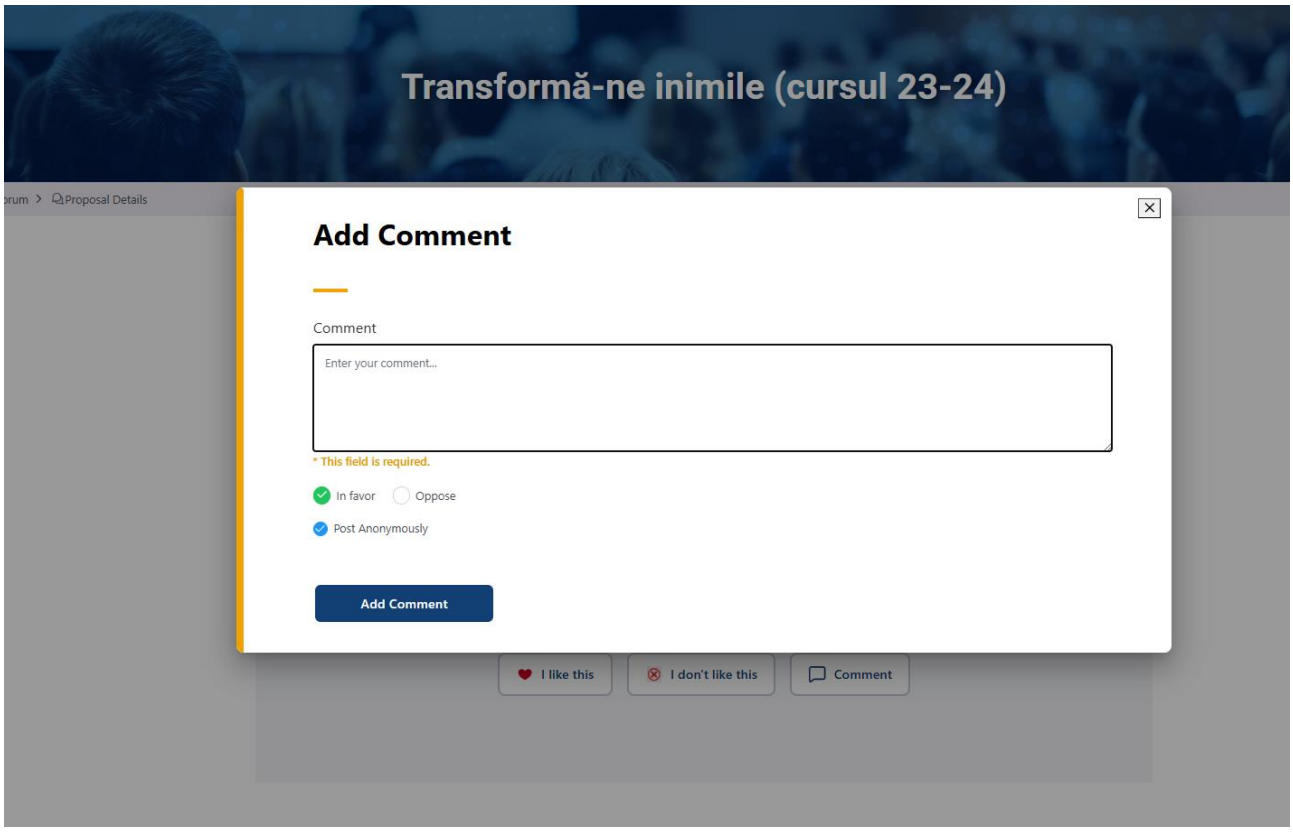


6.1.5 Argument Visualization and AI Tools

This section details how users interact with and contribute arguments to existing proposals, emphasizing the use of AI tools for content management and debate summarization.

1. Argumentation and Contribution

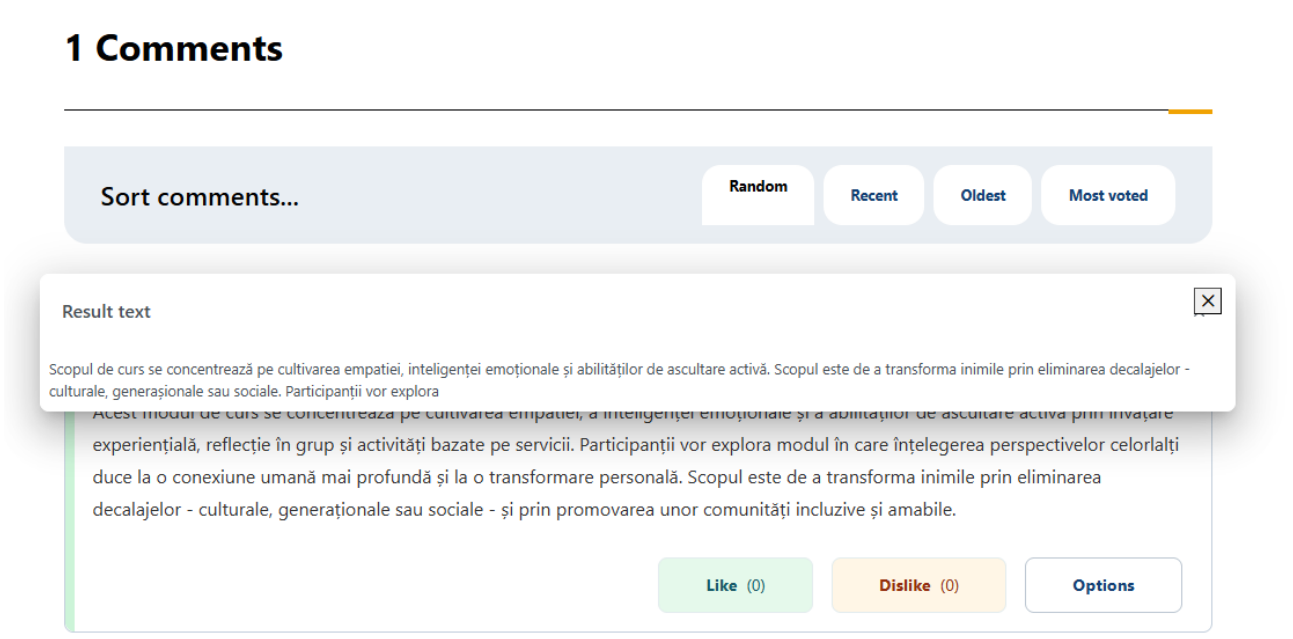
- **Access:** After selecting a proposal, users can view existing arguments categorized as "Pro" or "Con" (Supporting or Challenging).
- **New Arguments:** Users can submit new arguments, providing their stance (Position) and text content.
- **Argument Moderation:** Similar to proposals, all new arguments are processed by the AI toxicity evaluation tool. Arguments flagged as toxic are not immediately shown but are instead sent to a Human-in-the-Loop (HITL) system for moderator review.



2. AI-Powered Tools

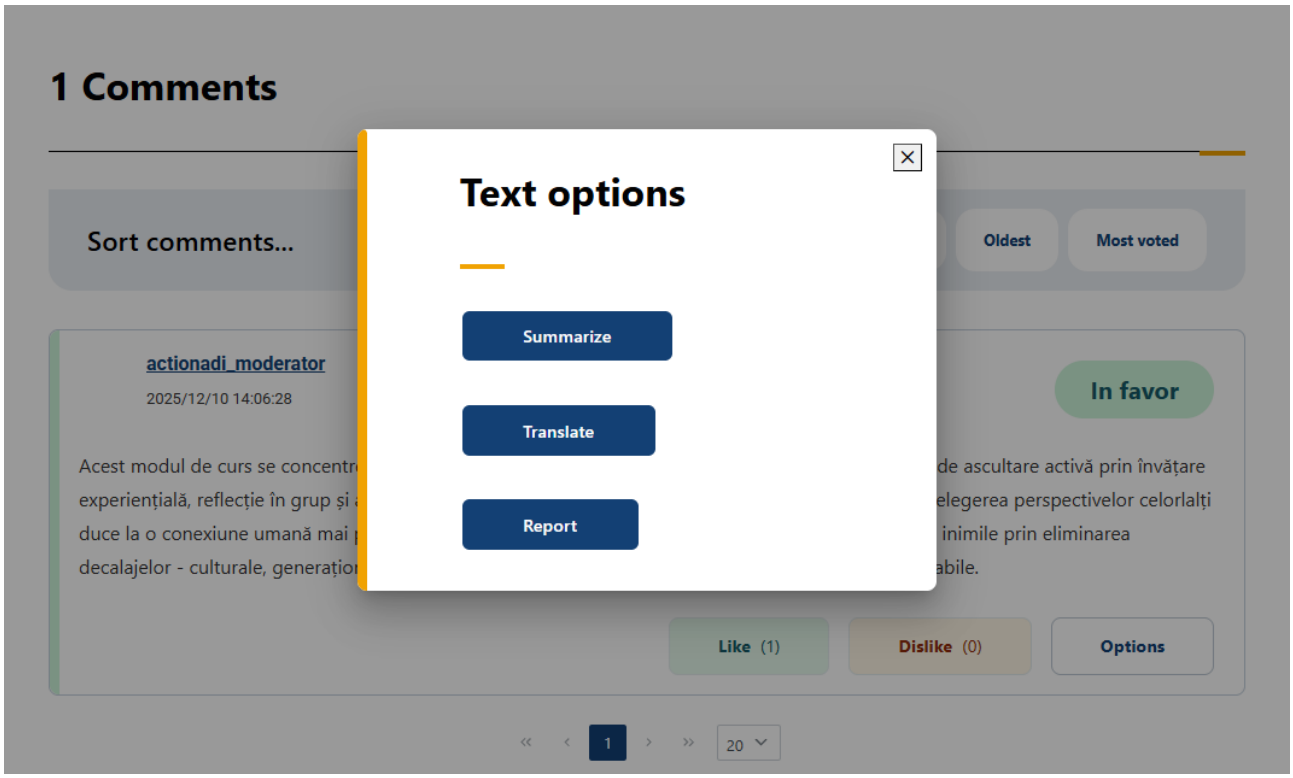
The platform provides several integrated AI services to enhance debate accessibility and clarity:

- **Toxicity Check:** Automatically evaluates all user-submitted text (proposals and arguments) for harmful or inappropriate content.
- **Summarization:** The accumulated arguments and discussion under a proposal can be summarized by the AI to provide users with a quick overview of the key points raised by the community. This summary is accessible via the **"Show Summarization"** button.
- **Translation:** Users can use a built-in translation tool to read or submit content in their preferred language, ensuring language is not a barrier to participation.



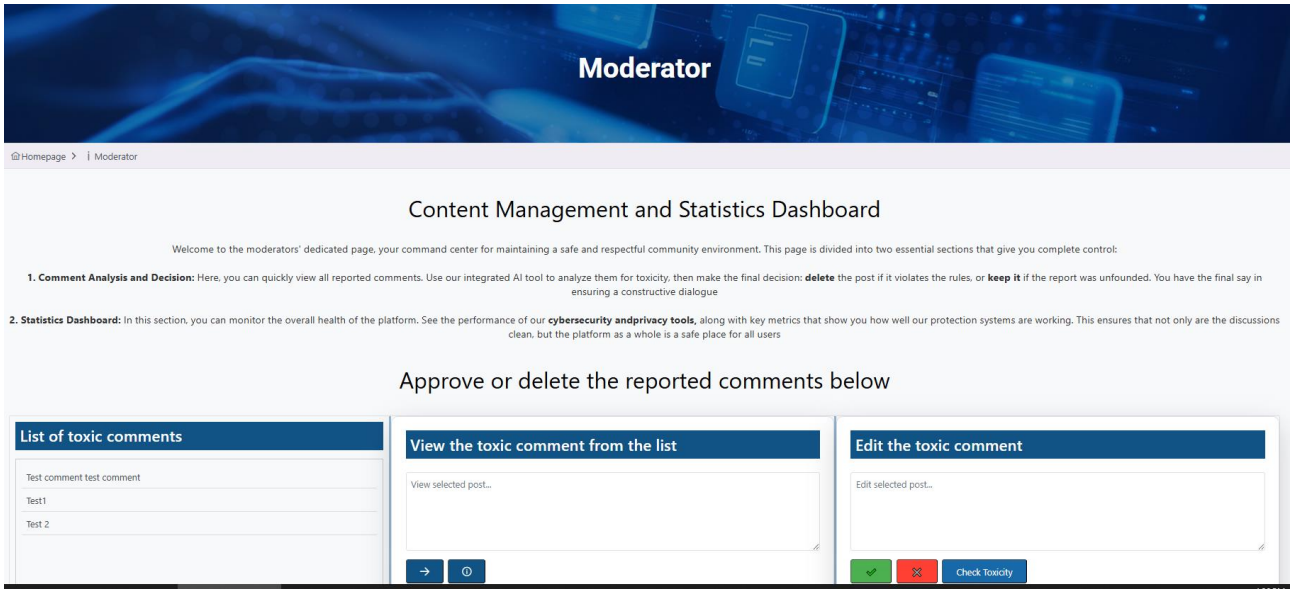
3. Argument Visualization and Voting

- **Voting:** Arguments can be individually upvoted or downvoted by users, helping to rank and elevate the most compelling points.
- **Reporting:** Users can report arguments as inappropriate, which also triggers a review by a human moderator.



6.1.6 Content Toxicity Check and Moderation

This section details the critical AI-powered process used to maintain a civil and respectful environment by automatically evaluating user-submitted content.



1. Real-Time Toxicity Evaluation

- **Mechanism:** When a user attempts to submit content (such as a proposal or an argument) containing potentially harmful or toxic language, the system intercepts the submission and sends the text to the AI toxicity evaluation tool.
- **Immediate Feedback:** If the text is flagged as toxic, the user is immediately notified with an **error message** (e.g., "The comment/proposal contains toxic language and cannot be submitted").
- **User Recourse:** The user is prevented from submitting the content in its current form but is given the option to modify the text to remove the toxic elements and reattempt submission.

2. Moderation Workflow

- **Purpose:** This tool serves as the first line of defense, ensuring that only civil content enters the main discussion threads, thereby promoting respectful engagement as required by the platform's rules.

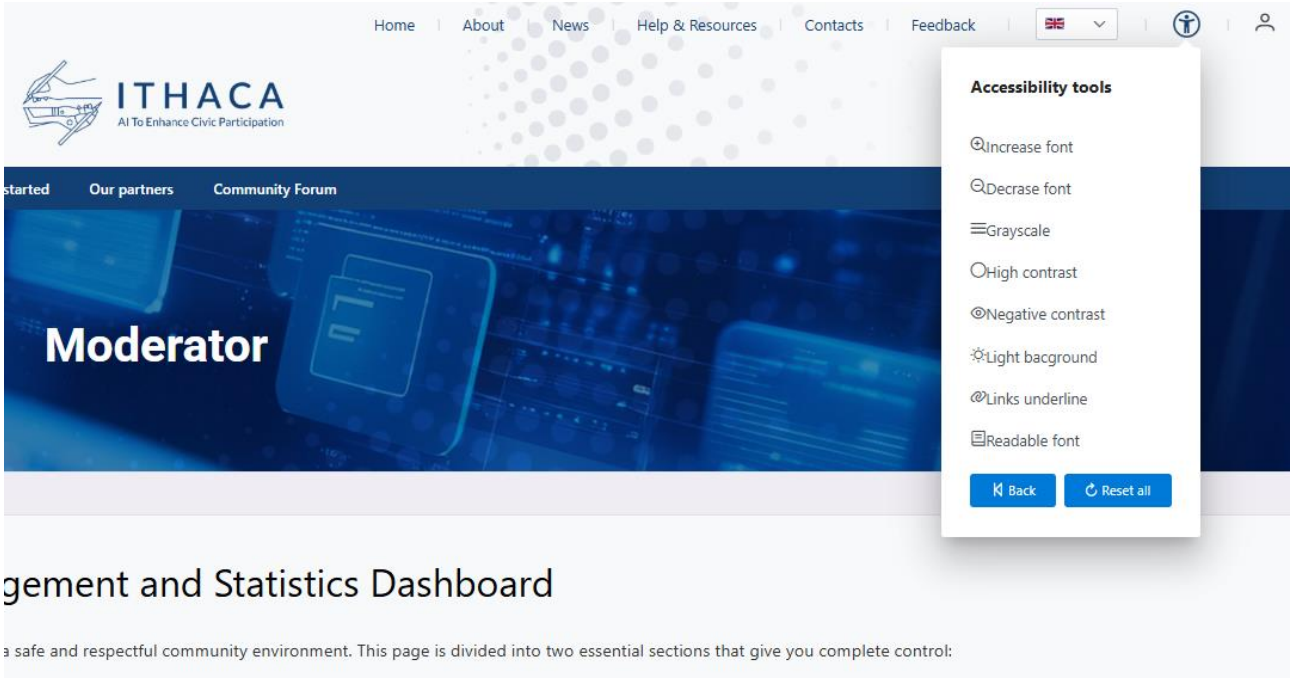
6.1.7 Accessibility Tools and Preferences

The platform integrates several tools to assist users with accessibility needs, honoring the preferences set during the registration process.

1. Accessibility Customization

- **Settings Toggle:** Users can access the platform's accessibility features via a dedicated icon.

- **Text Size:** The interface allows users to increase or decrease the text size, making content easier to read for those with visual impairments.
- **High Contrast:** A high-contrast color scheme can be toggled on to improve readability and visibility of text and interactive elements.
- **Language Selection:** The preferred display language can be changed from the accessibility menu (e.g., between English, Romanian, and Slovak).



2. Voice Input and Output

- **Voice Recognition (Input):** Users can click a dedicated microphone icon to enable voice-to-text input, allowing them to dictate comments, proposals, or search queries instead of typing. This feature supports users with motor disabilities.
- **Text-to-Speech (Output):** The platform features a speech service (speaker icon) that reads interface elements and textual content aloud, providing support for visually impaired users.

3. Platform Integration

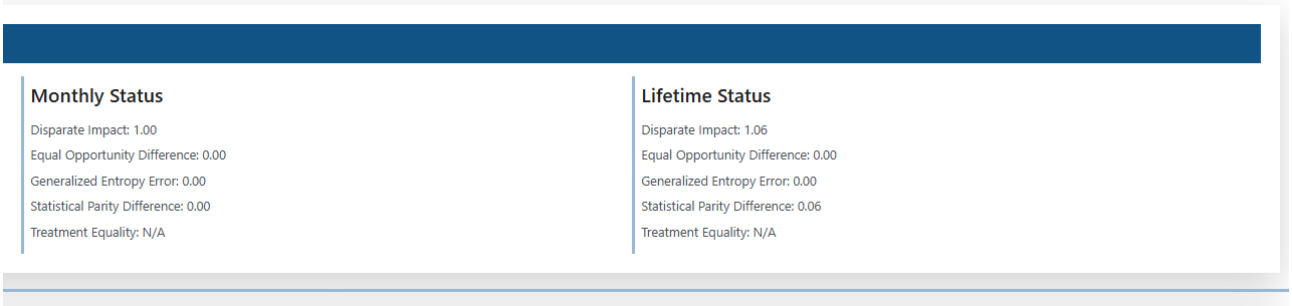
- **Persistence:** The accessibility and language choices made by the user are saved to their profile, ensuring a consistent experience across sessions and devices.

6.1.8 Ensuring AI Integrity and Fairness

This section outlines the platform's commitment to ethical AI use, focusing on transparency and user control over automated services.

1. AI Fairness and Transparency

- **AI Fairness Button:** A dedicated "AI Fairness" button or link is available on the platform home screen in the Smart Tools(AI features) area.
- **Transparency Dashboard:** Clicking this button leads the user to a dedicated dashboard or information page explaining how the AI models work, what data they use, and what steps are taken to mitigate bias in their operation (e.g., in content moderation and summarization).
- **Integrity:** The system monitors AI tool usage to ensure integrity and prevent misuse.



2. User Control and Feedback

- **Control over AI Tools:** The platform gives users control over certain aspects of the AI tools, honoring the principle of keeping users at the center of the technology.
- **Feedback Mechanism:** Users are provided with avenues to submit feedback regarding the performance or perceived bias of the AI tools, contributing to continuous improvement and fairness.

6.2 Operation in pilot contexts

The operation of the ITHACA platform in the pilot contexts of **Brasov** (Romania) and **Martin** (Slovakia) played a central role in validating the platform under real-world conditions and in guiding its iterative refinement. Rather than serving solely as a deployment exercise, the pilot operation functioned as a **feedback-driven validation cycle**, directly influencing technical requirements, system behaviour, and user-facing functionalities described in earlier chapters of this deliverable.

During the pilot operation, the platform was used by citizens, moderators, and municipal stakeholders to participate in structured online discussions, submit proposals, interact with AI-supported tools, and explore argument visualisations. Continuous monitoring of platform usage, combined with qualitative and quantitative evaluation activities (as reported in **D4.3**), enabled the consortium to identify strengths, limitations, and areas requiring adaptation.

Several **key adjustments implemented during or after pilot operation** can be directly linked to the evaluation outcomes:

- **User workflows and navigation** were refined following observations that first-time users experienced difficulties in locating core functionalities. As a result, simplified navigation paths, clearer action prompts, and improved onboarding mechanisms were introduced, building on the usability requirements reviewed in Chapter 3.

- **AI-supported functionalities**, particularly summarisation and moderation tools, were fine-tuned based on pilot feedback indicating the need for greater transparency and balanced representation of viewpoints. This led to adjustments in AI configuration, clearer explanatory messages, and reinforced human-in-the-loop mechanisms, as reflected in the final technical requirements and system architecture.

- **Gamification mechanisms** were revised after pilot participants reported uncertainty regarding reward logic and fairness. In response, mission definitions, scoring criteria, and feedback mechanisms were clarified to better align incentives with meaningful participation rather than activity volume alone.

- **Performance and stability improvements** were introduced as part of pilot operation, informed by real usage patterns and complemented by targeted stress and load testing. These improvements ensured that the platform could support peak participation periods without degradation, linking operational experience with the performance requirements described in Chapter 3 and the testing activities reported in Chapter 7.

Operational differences between the two pilot sites were also taken into account. While both cities followed the same core platform setup, local contextual factors—such as participation culture, moderation practices, and language—highlighted the importance of **configurability and flexibility**, reinforcing the architectural choices described in Chapter 4.

Overall, the operation of the platform in Brasov and Martin confirmed the value of an **iterative, evaluation-informed development approach**. The pilot contexts did not merely validate a predefined solution but actively shaped the final version of the platform, ensuring that the ITHACA system presented in this deliverable reflects **real user needs, operational constraints, and validated improvements**, rather than theoretical assumptions.

7. Testing and Validation

7.1 User Acceptance Testing

7.1.1 Evaluation of the Gamification System

The evaluation of the Gamification system integrated into the ITHACA platform aimed to assess the users' experience, contentment, as well as whether it influences active engagement with civic issues through instigating intrinsic motivation. This process was also designed to test the functionality of the underlying game mechanics including missions, XP points & level system, badges, etc.
The evaluation procedure took place at the Electrical and Computer Engineering (ECE) Department of the University of Patras. Nine students (undergraduate, BSc, and PhD ones) of the ECE Departement, aged 24-29, participated in the evaluation in-person. The participants were adequately familiar with gamification as well as forum-like systems.

As designed by CERTH, the assessment plan included a set of questionnaires and surveys that the participants were asked to fill out:

- Pre-questionnaire: The participants were asked to express their sentiments and expectations before the procedure.
- Micro-survey: A short survey that the participants had to fill out after completing a mission to record their feelings and progress in each type of user engagement level.
- Mid-mini survey: Participants were asked to describe their experience using the gamification system.
- Post-Questionnaire: Participant final thoughts and sentiments were collected through this questionnaire at the end of the evaluation process.

At the beginning of the procedure, a facilitator was guiding the participants through the missions of the gamification tool, and after completing the mid-mini survey session, the participants were free to explore the capabilities of the gamification system while navigating the ITHACA platform. The plan followed by the facilitator during the assessment is outlined in Table 4.

Table 4

Block	What happens
A. Welcome & consent	Brief purpose, privacy; sign consent; hand out accounts
B. Pre-questionnaire	Baseline motivation & expectations
C. Guided tasks (think-aloud)	Step-by-step missions + micro-surveys after each reward
D. Mid mini-survey	Flow & quick UX clarity check
E. Open exploration	Free play: climb rank, earn a badge
F. Post-questionnaire	GAMEX + gamification-UI usability
G. Debrief (group)	What worked, what didn't; suggestions
H. Wrap-up	Incentives, thanks

Result summary:

Positive Outcomes:

- Around 60% of the participants showed excitement in participating in the whole process, while the rest of the participants expressed positive emotions throughout the evaluation procedure. The latter indicates that the design of the gamification tool, along with the mission structure and badge system, achieves to incite interest and curiosity among participants, which ultimately led them to successfully engage with the platform's civic participation procedures.
- All participants completed missions quickly and did not have any major problems with handling the gamification tool (e.g., trouble finding or achieving missions, badges, etc.)
- Top Motives that motivated and aided the participants to enjoy the whole session: i) the fun sentiment of completing a mission, ii) the noble competitive spirit, and the urge to relate with friends/acquaintances while completing missions, iii) the curious nature of the participants who wanted to explore all gamification and platform features, iv) the participants communicated with one another to complete missions (even though no one achieved a "Relatedness" mission, they related with one another over completing missions).
- The participants interacted satisfactorily on various topics in the "Proposals" section, meaning that the gamification system enhanced civic engagement and discussion over civic issues.

Critical Outcomes:

- Implementation weaknesses were detected throughout the procedure, such as some missions not having the expected functionality or rewards (experience points, badges), while participants observed that badges, progress feedback, and the leaderboard were not fully informative and/or understandable.
- The participants' trust and interest in the gamification mechanism were compromised towards the end of the session due to these weaknesses, which also led to feelings of confusion.

More details regarding the evaluation of the gamification module and the collected outcomes are thoroughly outlined in Chapter 4.2.3 of D4.3. In this Chapter, a detailed table (Table 4) containing issues to be addressed based on how critical they are for the gamification system to operate smoothly and be optimized before wider deployment, is included. Critical issues such as fixing mission design and reward system weaknesses as well as optimizing progress feedback mechanisms, were given priority, to enhance gamified civic engagement.

Overall, the participants engaged with great interest in both the gamification system and the platform. Due to their engineering background, they quickly became familiar with the gamification system's functionality, which led to the fast completion of the evaluation process (<1 and a half hours). It is worth noting that during the evaluation of the gamification module, they quickly discovered and completed most of the guided tasks (even before the facilitator told them to do so), and then they proceeded to complete more missions, other than the ones the facilitator indicated. In parallel with their engagement with the gamification system, they started engaging more with the rest of the platform's functionalities and modules, which ultimately led them to detect missing functionalities and bugs. Therefore, the evaluation of the gamification engine transformed into platform user testing.

7.2 Pilot testing results

Based on the findings from the pilot evaluations, the following recommendations were made for improving the ITHACA platform:

1. **Usability and Navigation**:
    - Simplify the platform's **navigation structure** to reduce **information overload** and make key actions (e.g., posting, reading summaries, reporting) more intuitive.
    - Enhance **first-time user onboarding** to ensure smoother entry and understanding of core features, especially for **vulnerable groups** or those with limited digital literacy.
2. **AI Transparency and Fairness**:
    - Provide **clearer explanations** and **contextual information** alongside AI-generated content to improve **trust** and **transparency**. For example, displaying reasons for **toxicity flags** or offering links to relevant **moderation policies** could help users understand AI decisions better.
    - Ensure **minority views** are better represented in **AI summaries**, possibly by adding a mechanism to highlight **underrepresented opinions**.
3. **Gamification**:
    - Revise the **gamification system** to make scoring and ranking rules **clearer** and more aligned with **meaningful contributions** rather than activity alone.
    - Address **technical instabilities** within the gamification module to ensure smooth tracking of user progress and rewards.
4. **Performance and Scalability**:
    - Continue monitoring system performance and scalability, especially as user load increases. Any **performance bottlenecks** or **system failures** must be addressed promptly to maintain user trust and satisfaction.
5. **Accessibility Improvements**:
    - Make **accessibility features** more visible and easily discoverable by incorporating a **dedicated accessibility menu** or **tutorial** for first-time users.

Overall, the pilot phase in **Brasov** and **Martin** provided valuable insights into how the **ITHACA platform** functions in real-world settings. While the platform demonstrated strong performance in terms of usability, AI functionality, and engagement, improvements in **accessibility**, **AI transparency**, and **gamification fairness** are necessary to ensure that the platform is effective for all users, especially those from vulnerable or underrepresented groups. The recommendations outlined above should help enhance the platform's **usability**, **trustworthiness**, and **long-term sustainability**. More information can be found on D4.3.

8. Challenges, Lessons Learned, and Mitigations

The development and deployment of the **ITHACA platform** encountered various challenges throughout its lifecycle, from technical hurdles to ethical concerns. In this chapter, we explore the key challenges faced during the platform's design, development, and pilot deployment phases in **Brasov** and **Martin**. Alongside these challenges, we reflect on the lessons learned and outline the **mitigation strategies** implemented to overcome these difficulties and ensure the platform's continuous improvement.

8.1 Technical challenges

The **technical challenges** faced during the development of the **ITHACA platform** primarily stemmed from the complexity of integrating multiple advanced technologies and ensuring that the system operated efficiently across different environments. Key technical challenges included:

1. **System Performance and Scalability**:
    - **Challenge**: During the pilot phase, ensuring that the platform could handle the expected **user load** and deliver **consistent performance** across both **Brasov** and **Martin** was a significant concern. While **load testing** revealed acceptable performance under typical usage, there were still minor **glitches** and **delays** during peak periods.
    - **Lesson Learned**: **Performance monitoring** should be an ongoing process, and it's critical to continuously optimize system resources, particularly when user demand increases unexpectedly.
    - **Mitigation Strategy**: To address these challenges, the **ITHACA platform** incorporated **scalable cloud infrastructure**, with **auto-scaling** features, and optimized performance by reviewing bottlenecks in the **back-end services**. Furthermore, **stress testing** was used during the pilot to ensure resilience under extreme conditions, while **load balancing** and **distributed services** were explored for future scalability.
2. **Integration of AI Components**:
    - **Challenge**: Integrating AI-driven components such as **summarization**, **toxicity detection**, and **fairness checks** posed challenges due to the complexity of ensuring their **accuracy** and **reliability** in real-world scenarios. Some **AI-generated summaries** were perceived as overly simplified, and the **toxicity detection** tool occasionally flagged content inaccurately.

- **Lesson Learned**: **AI models** need constant **fine-tuning** to ensure they can handle the nuances of **real-world data** and reflect diverse perspectives in discussions.

- **Mitigation Strategy**: The platform incorporated **human-in-the-loop** mechanisms where **moderators** could intervene and correct AI-generated outputs. Additionally, the **AI components** underwent further training using **real-world data** from pilot discussions to improve performance and reduce bias.

3. **Security and Privacy Concerns**:

   - **Challenge**: Ensuring that the platform adhered to strict **data protection** and **security** standards while managing sensitive information was a critical challenge, especially given the **GDPR** compliance requirements.

   - **Lesson Learned**: **Data minimization** and **anonymization** protocols must be rigorously implemented to protect user privacy.

   - **Mitigation Strategy**: The platform adopted **end-to-end encryption** and **anonymization** techniques to ensure the secure handling of user data. **Role-based access control** (RBAC) was implemented to ensure that only authorized personnel had access to sensitive information, and regular **security audits** were conducted to identify potential vulnerabilities.

8.2 Ethical and legal challenges

The **ethical and legal challenges** in developing the ITHACA platform primarily revolved around ensuring that the platform's technologies aligned with European **ethical standards**, including **fairness**, **transparency**, and **inclusivity**. Specific challenges included:

1. **AI Fairness and Bias**:

   - **Challenge**: One of the primary ethical concerns was ensuring that the **AI tools** did not inadvertently reinforce existing biases or undermine the fairness of online civic discussions. AI-generated summaries sometimes excluded minority perspectives, and **toxicity detection** occasionally flagged content inaccurately, raising concerns about fairness.

   - **Lesson Learned**: Ensuring **AI transparency** and **accountability** is crucial to maintaining **trust** in AI systems.

   - **Mitigation Strategy**: To mitigate these issues, the platform implemented regular reviews of AI performance, ensuring that AI outputs were regularly audited for **bias** and **fairness**. Furthermore, **human-in-the-loop** moderation mechanisms were integrated to allow human intervention when AI made biased or incorrect

assessments.

2. **Data Privacy and GDPR Compliance**:

    - **Challenge**: Given the nature of the platform, managing **personal data** and ensuring compliance with **GDPR** regulations was a significant challenge. Participants' data, particularly in relation to **user-generated content**, had to be handled carefully to avoid privacy violations.

    - **Lesson Learned**: Strict **data governance** and **privacy safeguards** are essential in any civic participation platform.

    - **Mitigation Strategy**: The platform ensured compliance by adopting **privacy-by-design principles** and integrating a robust **Personal Information Management System (PIMS)**. This allowed users to manage their data, revoke consent, and control the use of their personal information.

3. **Ensuring Inclusivity in Civic Participation**:

    - **Challenge**: Ensuring that the platform was accessible to all citizens, including those from vulnerable or marginalized groups, was a complex issue. Accessibility features needed to cater to a wide range of needs, from individuals with disabilities to people with low digital literacy.

    - **Lesson Learned**: **Inclusive design** must be a core principle throughout the platform's development.

    - **Mitigation Strategy**: To address this, the platform implemented a range of **accessibility features**, such as high-contrast modes, text-to-speech, and **multilingual support**. Moreover, **user testing** with vulnerable groups was conducted to identify any potential barriers to participation.

8.3 Usability and accessibility challenges

The **usability and accessibility** challenges centered around ensuring that the **ITHACA platform** was user-friendly, intuitive, and accessible to a diverse range of users, including those with varying levels of **digital literacy** and **physical abilities**. Specific challenges included:

1. **Navigational Complexity**:

    - **Challenge**: Users, especially those with lower digital literacy, struggled to navigate the platform efficiently. Some found the **interface** to be overwhelming and complex, with too much information displayed at once.

- **Lesson Learned**: Platforms must prioritize **simplicity** in design, especially when targeting a wide range of users.

- **Mitigation Strategy**: The platform's **user interface** was simplified by reducing unnecessary complexity and ensuring that key actions, like **posting** or **reporting**, were easily discoverable. **Guided onboarding** was implemented to help users familiarize themselves with the platform's features step by step.

2. **Visibility of Accessibility Features**:

   - **Challenge**: Although accessibility features were integrated into the platform, users often found it difficult to locate and utilize them.

   - **Lesson Learned**: **Accessibility options** must be both **visible** and **easy to use**.

   - **Mitigation Strategy**: To make accessibility features more prominent, the platform introduced a **dedicated accessibility menu** and included **tutorials** to guide users in enabling accessibility tools. Additionally, the platform ensured that these features were consistent across different pages.

3. **Mobile Usability**:

   - **Challenge**: Many users in the pilot sites accessed the platform via **mobile devices**, and the mobile version of the platform was initially not optimized, leading to a poor user experience.

   - **Lesson Learned**: **Mobile optimization** is essential for ensuring wide accessibility.

   - **Mitigation Strategy**: The platform was **re-designed** to be **mobile-friendly**, ensuring that users had a seamless experience across both desktop and mobile devices.

8.4 Mitigation strategies

In response to the challenges identified in the previous sections, several **mitigation strategies** were adopted to improve the platform and ensure that it met its goals of **usability**, **accessibility**, **fairness**, and **privacy**. These strategies included:

1. **Continuous User Feedback and Iteration**:

   - **Strategy**: Incorporating **continuous feedback loops** from users in **Brasov** and **Martin** allowed the development team to refine and improve the platform. Pilot testing, surveys, and focus groups provided valuable insights into user preferences

and pain points, which informed ongoing development.

2. **AI Transparency and Human Oversight**:
    - **Strategy**: To address concerns about AI bias and transparency, the platform integrated **human-in-the-loop** mechanisms, ensuring that moderators could intervene in AI decisions. Regular audits of AI outputs were conducted to ensure that AI systems were fair and accurate.
3. **Enhanced Data Privacy Measures**:
    - **Strategy**: **GDPR compliance** was ensured through the implementation of **privacy-by-design** features, including anonymization, **data consent management**, and transparent privacy policies. This allowed users to feel secure in their participation and helped mitigate concerns about data misuse.
4. **Simplified User Interface**:
    - **Strategy**: The user interface was **simplified** to enhance usability, particularly for those with **low digital literacy**. This included **reduced clutter**, clearer navigation paths, and **helpful onboarding tutorials**.
5. **Accessibility and Inclusivity Focus**:
    - **Strategy**: The platform was designed with **accessibility** at its core, including features like **speech-to-text**, **multilingual support**, and **keyboard navigation**. These features were made more visible and accessible to users, and feedback from users with disabilities was incorporated into ongoing improvements.

The challenges faced during the development and deployment of the **ITHACA platform** were significant but manageable through a combination of **technical innovation**, **ethical consideration**, and a strong **focus on user feedback**. The lessons learned have been instrumental in refining the platform's **usability**, **AI functionality**, and **accessibility** features, ensuring that **ITHACA** is better positioned to support **inclusive**, **transparent**, and **fair** civic engagement in the future. By continuing to address these challenges, ITHACA can serve as a model for democratic innovation and AI-enhanced participation.

9. Conclusions and Future Work

In this chapter, we reflect on the overall progress and impact of the **ITHACA platform** based on its development and evaluation phases, particularly in relation to the **GA (Grant Agreement) objectives**. We also discuss the **sustainability** and **exploitation perspectives** of the platform, providing insights into how it will continue to evolve and be utilized after the project's conclusion. Lastly, we offer **recommendations for future development** to ensure that **ITHACA** remains a relevant and effective tool for **civic participation** in the future.

9.1 Alignment with GA objectives

The **ITHACA project** was designed to enhance **civic participation** through **AI-driven technologies** that foster transparency, inclusivity, and democratic engagement. Over the course of the project, several key objectives outlined in the **Grant Agreement (GA)** were met, demonstrating that **ITHACA** is on track to achieve its intended impact. These objectives include:

1. **Enhancing Civic Engagement and Inclusivity**:
    - **Objective**: One of the primary goals of **ITHACA** was to create a platform that enables citizens—particularly those from vulnerable and underrepresented groups—to participate in democratic processes.
    - **Achievement**: Through the pilot deployments in **Brasov** and **Martin**, the platform demonstrated its ability to facilitate **active civic participation**, offering tools such as **AI-based summarization**, **toxicity detection**, and **argument visualization** to help users engage more effectively in local governance.
2. **AI-Driven Support for Deliberative Democracy**:
    - **Objective**: The project aimed to leverage **AI** to support more transparent, fair, and efficient democratic decision-making processes.
    - **Achievement**: The **AI components**, such as **summarization** and **toxicity detection**, were successfully integrated into the platform, offering citizens and moderators tools to navigate and manage large volumes of information. Additionally, the **algorithmic impact assessment** (AIA) ensured that the platform’s AI tools were fair, transparent, and aligned with ethical standards.
3. **Data Privacy and Ethical Considerations**:
    - **Objective**: Ensuring that the platform adhered to **GDPR** standards and other **ethical guidelines** was a key objective.
    - **Achievement**: **ITHACA** integrated robust **privacy protection** mechanisms, including **data anonymization**, **user consent management**, and secure data

handling, aligning with EU privacy regulations and the **Ethical Guidelines for AI**.

4. **Scalable and Modular Architecture**:

    - **Objective**: Another key GA objective was the creation of a **scalable and flexible architecture** capable of supporting the platform's expansion and adaptation.

    - **Achievement**: The **hybrid architecture** (combining microservices with event-driven models) implemented in **D3.1** and refined in **D3.3** ensures that **ITHACA** can be easily adapted to different civic contexts and scales of engagement, making it a robust solution for future expansion.

Overall, the project successfully met the **GA objectives** related to technology development, user engagement, ethical standards, and scalability. The alignment with the **EU's Horizon Europe** priorities for **inclusive democracy** and **AI-driven innovation** positions **ITHACA** as a model for future civic engagement platforms.

9.2 Sustainability and exploitation perspectives

The **sustainability** and **exploitation** of the **ITHACA platform** are key to ensuring that the benefits of the project continue beyond its formal conclusion. The **Exploitation Plan** outlines how the platform will be maintained, developed, and deployed for long-term impact. To learn more about the Exploitation of the platform, you can read D6.3 Exploitation Plan.

9.3 Recommendations for future development

To ensure that **ITHACA** remains an effective tool for **democratic participation** and continues to evolve to meet the needs of users, the following **recommendations** are provided for future development:

1. **Refine AI Models for Fairness and Coverage**:

    - While the platform has made great strides in **AI transparency** and **fairness**, further efforts are needed to refine the **summarization algorithms** to better represent **minority viewpoints** and ensure that AI-generated content does not under-represent any group.

    - **Recommendation**: Invest in **diverse training datasets** to ensure that the AI models can handle complex and nuanced citizen contributions, particularly those from marginalized or underrepresented communities.

2. **Simplify Onboarding and Enhance Usability**:

    - Although the platform's **user interface** has been simplified, further improvements are needed to ensure a **seamless user experience**, particularly for **new users** and **vulnerable groups**.

    - **Recommendation**: Continue simplifying the **onboarding process**, adding more **interactive tutorials** and ensuring that the **navigation** is intuitive for all users, especially those with **low digital literacy**.

3. **Strengthen Integration with Other Platforms**:

    - To broaden the platform's reach and utility, **ITHACA** could benefit from integrating with **external systems**, such as **municipal management platforms**, **social media**, or **local government portals**.

    - **Recommendation**: Develop **API integrations** with other civic engagement and governance platforms to allow for **cross-platform data sharing**, creating a more interconnected civic ecosystem.

4. **Expand International Reach and Multi-Language Support**:

    - The platform currently supports **multilingual capabilities** in **Brasov** and **Martin**. However, to broaden its impact, future iterations should include support for even more languages and be adaptable to different **cultural contexts**.

    - **Recommendation**: Focus on expanding the platform's **language support** and **localization** features to ensure **global scalability** and inclusivity in more diverse regions.

5. **Maintain Strong Governance and Ethical Oversight**:

    - Given the **ethical considerations** surrounding the use of **AI** in civic participation, ongoing governance mechanisms are needed to ensure the platform remains aligned with **EU regulations** and **ethical standards**.

    - **Recommendation**: Establish an **ethical oversight committee** for the continuous evaluation of **AI fairness**, **data privacy**, and **transparency** to ensure compliance with **GDPR** and **AI ethics guidelines**.

The **ITHACA platform** has made remarkable strides in its mission to **enhance civic participation** by integrating cutting-edge **AI technologies** aimed at improving transparency, inclusivity, and the efficiency of democratic engagement. Through the iterative development and deployment phases, particularly in the **pilot contexts** of **Brasov** and **Martin**, the platform has proven its ability to

empower citizens—especially those from vulnerable and underrepresented groups—by providing them with innovative tools to participate actively in local governance.

The platform's **AI-driven features**, including **summarization tools**, **toxicity detection**, and **argument visualization**, have contributed to streamlining the process of civic engagement. These features, along with **gamification elements** and enhanced **user workflows**, have increased **user engagement**, especially in deliberative democratic processes. However, feedback from the pilot phases has highlighted areas for **improvement**, such as the need for better **AI fairness**, **more transparent gamification systems**, and **simplified navigation** to ensure accessibility for all users, including those with limited digital literacy.

In alignment with the **GA objectives**, ITHACA has successfully met its goals in creating an **inclusive** and **transparent platform** that supports **deliberative democracy** through AI-enhanced tools. Additionally, the platform has adhered to **ethical** and **legal standards**, particularly with respect to **GDPR compliance**, **AI fairness**, and **data privacy**.

Looking forward, the platform is set to evolve with an emphasis on ensuring **sustainability** and **exploitation** through scalable, adaptable solutions that can be utilized by municipalities, public institutions, and other stakeholders. Key recommendations for **future development** include further refining **AI models** for enhanced **coverage and fairness**, optimizing **user onboarding** processes, and expanding the platform's reach internationally with more **multilingual support**. Additionally, integrating **cross-platform interoperability** and maintaining strong **ethical oversight** will be essential to ensure that ITHACA remains an inclusive and trusted tool for future democratic innovations.

In conclusion, the ITHACA platform has laid a strong foundation for improving **democratic participation** through **AI technologies**, and its future success hinges on continued development, feedback integration, and expanding its influence as a model for **inclusive governance** in the digital age.