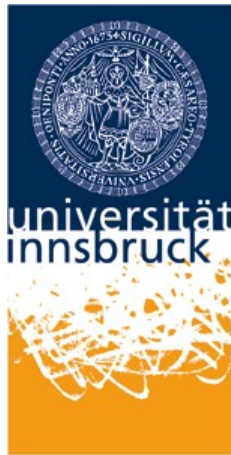


Leopold-Franzens Universität Innsbruck
Fakultät für Mathematik, Informatik und Physik

Institut für Informatik

Semantic Technology Institute Innsbruck



Bachelorarbeit

zur Erreichung des akademischen Grades

Bachelor of Science

Developing Online Game for Guessing Photo Age

von

Hasan Yücedag
(Matr.-Nr.: 12201941)

Submission Date: April 11, 2025

Supervisor: Adam Jatowt

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt durch meine eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Stellen, die wörtlich oder inhaltlich den angegebenen Quellen entnommen wurden, sind als solche kenntlich gemacht.

Ich erkläre mich mit der Archivierung der vorliegenden Bachelorarbeit einverstanden.

Ort und Datum:

6020 Innsbruck, 11.04.2025

Unterschrift:

Hymezdag

Disclaimer on the Use of AI

In the preparation of this work, AI tools were used for specific purposes. ChatGPT was utilized for chart creation, while Gemini, Grok, and ChatGPT were used for paraphrasing and linguistic refinement. Consensus AI was employed for source identification.

However, the work itself was not generated by AI. The content, structure, and arguments were independently developed, with AI assistance limited to improving readability and presentation. All AI-assisted modifications were critically reviewed to ensure accuracy and originality.

Contents

List of Figures	II
Abstract	1
1 Introduction	2
1.1 Motivation	2
1.2 Problem Statement	3
1.3 Objectives	3
1.4 Structure of the Thesis	4
2 Related Work	5
2.1 Gamification in Web Applications	5
2.2 Human Perception of Image Aging	6
2.3 Similar Online Games and Applications	7
3 Approach	9
3.1 Concept and Game Mechanics	9
3.2 User Interaction and Scoring System	10
3.3 Data Collection and Image Selection	10
3.4 Challenges in Age Estimation	11
4 Methodology	12
4.1 Development Environment and Technologies Used	12
4.2 Frontend and Backend Implementation	13
4.3 User Interface and Experience Design	14
4.4 Data Infrastructure and Image Processing	15
4.4.1 Database Design	15
4.4.2 Image Selection and Storage	18
4.4.3 Image Processing and Game Integration	18
5 Server Deployment and Infrastructure	20
5.1 Server Environment	20
5.2 Software Setup	21
5.3 Database Configuration	21
5.4 Application Deployment	21
5.5 Running the Application	22
5.6 Service Activation	22

Contents

5.7	Updates and Maintenance	22
6	Game Modes	24
6.1	Guess the Age of a Single Photograph	24
6.1.1	Gameplay Mechanics	25
6.1.2	Scoring System	26
6.1.3	Feedback and Learning	27
6.1.4	Demo Mode	27
6.2	Compare Two Photographs: Which is Older?	27
6.2.1	Gameplay Mechanics	28
6.2.2	Scoring System	29
6.2.3	Feedback and Learning	30
6.2.4	Demo Mode	30
6.3	Ranking and Scoring Mechanism	31
6.3.1	Point Tracking	31
6.3.2	Leaderboard	31
6.3.3	Gamification Elements	32
6.3.4	Points Aggregation Across Game Modes	32
7	User Instructions and Cohort Management	34
7.1	User Instructions	34
7.1.1	Our Mission Tab	34
7.1.2	How to Play and Point System	35
7.2	Cohort Management	35
7.2.1	Participant Demographics	36
7.2.2	Participation Guidelines	36
8	Evaluation	38
8.1	Testing Setup	38
8.2	User Feedback and Experience	39
8.3	Accuracy of User Guesses	41
8.3.1	Total Accuracy	41
8.3.2	Breakdown by Age Groups	42
8.3.3	Decade-wise Difficulty Across Users	43
8.3.4	Breakdown by Gender	45
8.4	Engagement and Retention Metrics	47
9	Future Work and Extensions	49
9.1	Key Game Mode and Feature Enhancements	49
9.1.1	Decade Match Mode: Deepening Historical Understanding	49
9.1.2	Basic Hint System: Enhancing Accessibility	50
9.2	Focused Dataset Expansion: Quality over Quantity	50
9.2.1	Curated Themed Collections: Adding Depth and Variety	50
9.3	Mobile Accessibility: Reaching a Wider Audience	51
9.3.1	Responsive Web Design Improvements: A Practical First Step	51

Contents

10 Conclusion	52
Bibliography	53

List of Figures

6.1	An example of Guess the Year mode	25
6.2	An example of the Timeline Challenge mode	28
6.3	Live scoreboard with static and dynamic point rankings.	33
6.4	User's point breakdown by game mode.	33
8.1	User feedback ratings across five evaluation categories	39
8.2	Comparison of accuracy rates between game modes	41
8.3	Accuracy rates by age group for both game modes	42
8.4	Accuracy comparison between male and female participants	43
8.5	Number of correct and incorrect answers per decade by game mode .	44
8.6	Distribution of users by number of gameplay sessions	47
8.7	Cumulative percentage of users by gameplay frequency	47

Abstract

This thesis presents the design and development of an engaging web-based platform that allows users to interact with historical photographs through two entertaining game modes. The first game challenges users to guess the year a single photograph was taken, focusing on images from 1930 to 1999. The second game presents users with two photographs and asks them to determine which one is older. The objective of this project is to create an educational and entertaining experience that combines gamification with historical exploration.

The platform was developed using a user-centered approach, incorporating modern web technologies for seamless functionality and an intuitive interface. A carefully curated dataset of historical images was assembled, ensuring diversity and chronological accuracy. The game mechanics include scoring systems and feedback mechanisms to enhance user engagement. Challenges such as image selection, age estimation, and user interface design were addressed during the development process.

The system was evaluated through user testing, focusing on metrics such as user satisfaction, engagement, and the accuracy of guesses. Results demonstrated the platform's ability to captivate users while providing an enjoyable way to explore history. Future work includes expanding the image database, introducing additional game modes, leveraging machine learning for automated age estimation, and developing a mobile application version to reach a broader audience.

1 Introduction

1.1 Motivation

The internet has become a powerful medium for creating engaging and interactive experiences. Gamification, the integration of game-like elements into non-game contexts, has emerged as an effective way to entertain, educate, and engage users. Historical photographs hold a unique charm, offering a glimpse into the past and sparking curiosity about the lives, events, and cultures of earlier times. However, many people struggle to contextualize these images or estimate their age accurately. This project aims to bridge the gap between entertainment and education by creating an interactive website where users can guess the age of historical photographs or compare two images to determine which is older. The platform seeks to provide a fun and educational experience that fosters curiosity about history while leveraging gamification principles.

1.2 Problem Statement

Despite the abundance of historical photographs available online, there is a lack of interactive platforms that make exploring these images both entertaining and educational. Existing tools often focus on static presentations or require significant prior knowledge to engage with historical content meaningfully. There is a need for an accessible and enjoyable platform that allows users to interact with historical images in a way that challenges their perception of time and enhances their understanding of history.

1.3 Objectives

The primary objective of this thesis is to design and develop an entertaining web-based application that enables users to:

- Guess the year a historical photograph was taken (1930–1999).
- Compare two photographs and determine which one is older.

The platform will incorporate user-friendly interfaces, gamified features such as scoring systems, and curated datasets of historical images to ensure an engaging experience. Additionally, this project aims to evaluate the platform’s effectiveness through user testing, focusing on metrics such as user engagement, satisfaction, and accuracy of guesses.

1.4 Structure of the Thesis

This thesis is organized as follows:

- **Chapter 2** provides a review of related work, including gamification in web applications, human perception of image aging, similar online games, and existing image databases for old photographs.
- **Chapter 3** discusses the approach taken in designing the platform, including game mechanics, user interaction, data collection, and challenges in age estimation.
- **Chapter 4** describes the methodology used for development, covering the technologies employed for frontend and backend implementation, user interface design, database management, and image processing techniques.
- **Chapter 5** details the server environment, software setup, database configuration, application deployment, and service activation. It also includes updates and maintenance procedures.
- **Chapter 6** outlines the two main game modes—guessing the age of a single photograph and comparing two photographs—and explains the scoring mechanism used in both modes.
- **Chapter 7** details the instructions provided to users and how the participant cohort was managed throughout the study.
- **Chapter 8** evaluates the platform based on user feedback, accuracy of guesses, performance metrics, and optimization efforts.
- **Chapter 9** explores potential future work and extensions, such as additional game modes, machine learning integration for automated age estimation, expanding the image database, and developing a mobile app version.
- **Chapter 10** concludes the thesis by summarizing key findings and discussing their implications.

2 Related Work

This section explores prior research relevant to developing a website where users guess the ages of people in photos. It examines how gamification enhances user engagement in web-based systems, the cognitive and visual processes involved in perceiving age from images, and the design and mechanics of existing online games and applications with similar objectives. By analyzing these areas, this review provides valuable insights into the technological and design considerations necessary for creating an interactive and entertaining user experience.

2.1 Gamification in Web Applications

Gamification, the integration of game-like elements such as points, badges, and leaderboards into non-game contexts, has emerged as a powerful strategy to enhance user engagement in web applications. Research demonstrates its efficacy in fostering motivation, participation, and retention, particularly in educational and interactive environments [1]. For a website centered on age guessing, gamification offers a framework to transform a simple task into a compelling, repeatable activity. Studies such as those by Hamari [4] emphasize that gamified systems can significantly increase user interaction by introducing competitive and reward-based mechanics. For instance, a scoring system that awards points for accurate age guesses could incentivize users to refine their skills and return frequently. Similarly, leaderboards, as explored in Nicholson [7], promote a sense of community and rivalry, encouraging users to compare their performance with others. The systematic review by Seaborn

2 Related Work

and Fels [10] further supports the use of immediate feedback loops—such as displaying the difference between a user’s guess and the actual age—as a means to sustain engagement and provide an educational byproduct.

In educational web applications, gamification has proven particularly effective. Domínguez [2] illustrate how incorporating challenges and progress tracking in a learning platform enhances user satisfaction and persistence, a finding that aligns with the goals of the proposed website. By adapting these principles, the age guessing platform could implement tiered difficulty levels or badges for milestones (e.g., correctly guessing ages across diverse demographics), thereby deepening user investment. These insights suggest that a gamified approach could elevate the website from a mere novelty to a dynamic, user-centric experience.

2.2 Human Perception of Image Aging

The human ability to estimate age from photographs is a complex perceptual process, influenced by visual cues and contextual factors, and has been extensively studied in comparison to machine performance. Understanding these dynamics is critical for designing an age guessing website that leverages both human intuition and technological precision to create an entertaining contrast.

Research indicates that human age estimation is inherently variable, with accuracy affected by factors such as gender, facial expressions, and image quality [9]. For example, Han found that humans tend to overestimate ages when facial expressions convey fatigue or stress, while underestimating them in cases of youthful features like smooth skin [5]. This variability introduces an element of unpredictability that can enhance the entertainment value of the website, as users grapple with their own perceptual biases.

Comparative studies of human versus machine performance provide further insight. Guo and Mu demonstrate that machine learning models, particularly those em-

2 Related Work

playing convolutional neural networks (CNNs), consistently outperform humans in controlled settings, achieving mean absolute errors as low as 3-5 years on datasets like FG-NET and MORPH [3]. However, human guesses, while less precise, carry a social and intuitive appeal that machines cannot replicate. This dichotomy, as explored by Han, suggests a unique opportunity for the website: integrating a feature that juxtaposes user estimates with AI-generated predictions [5]. Such a comparison not only entertains but also educates users about the nuances of aging and perception.

Moreover, recent advancements, such as those reported by Shin [11], highlight the potential of relative age estimation—judging whether someone appears older or younger than a reference—as a more intuitive approach for humans. This could inform the website’s design by offering varied guessing modes, enhancing both accessibility and engagement. Collectively, these findings underscore the value of blending human perception with machine accuracy to create a distinctive and interactive user experience.

2.3 Similar Online Games and Applications

Several existing online games and applications provide practical precedents for designing an age guessing website, offering insights into user interface design, engagement strategies, and potential use cases. These platforms, while varying in scope, share the common goal of making age estimation an enjoyable activity, often through gamified elements and community features.

One prominent example is AgeGuessr (<https://www.ageguessr.com/>), a web-based game where users estimate the ages of celebrities based on photographs. The platform employs a straightforward scoring system, awarding points for guesses close to the actual age, and includes a leaderboard to foster competition. This model highlights the effectiveness of simplicity and immediate feedback in sustaining user interest, elements that could be directly adapted to our website. Similarly, GuessTheAge

2 Related Work

(<https://guesstheage.org/>) expands the scope by incorporating user-submitted photos, paired with AI-generated age estimates for comparison. Its inclusion of global leaderboards and social sharing options exemplifies how community-driven features can amplify engagement.

Beyond these digital platforms, an unexpected parallel emerges from offline applications, such as age guessing games available on marketplaces like Etsy (https://www.etsy.com/market/age_guessing_game). These printable games, often used at events like bridal showers or reunions, task participants with estimating ages from historical or personal photos. While not web-based, their popularity underscores a broader cultural appeal for age guessing as a social activity, suggesting potential for the website to target niche audiences or event-based contexts—perhaps through downloadable content or group play modes.

3 Approach

This chapter presents the methodology for developing "Guess the Age of Photos". The game engages users in estimating the acquisition dates of historical photographs, utilizing a subset of 10,150 images from the "Date Estimation in the Wild" (DEW) dataset by Müller, Springstein, and Ewerth (2017)[6]. The approach encompasses the conceptual framework, user interaction and scoring design, data selection process, and challenges inherent in age estimation, ensuring an integrated system that combines entertainment with technical precision.

3.1 Concept and Game Mechanics

The conceptual foundation of "Guess the Age of Photos" is to transform the task of dating historical photographs into an entertaining and educational experience through gamification. Leveraging the DEW dataset's Flickr images from 1930 to 1999, the game introduces two core modes—"Guess the Year" and "Timeline Challenge"—to engage users with varied mechanics. "Guess the Year" involves estimating the acquisition year of a single image, while "Timeline Challenge" requires comparing two images to determine which is older. These modes utilize the dataset's ground truth years (GT) to assess accuracy, fostering historical exploration within a competitive framework. The approach emphasizes randomization and a dynamic difficulty structure, with further mechanics detailed in Chapter 6.

3.2 User Interaction and Scoring System

User interaction is designed to be intuitive and motivating, laying the groundwork for a seamless experience. The platform offers a demo mode for unregistered users and requires registration (name, age, gender) for full access, directing players to a landing page for mode selection. Interaction involves submitting guesses—numeric for "Guess the Year," binary for "Timeline Challenge"—with immediate feedback on accuracy and scores. For accurate predictions and 5-year interval predictions, the scoring system awards a maximum of 10 points, with dynamic adjustments based on error margins or temporal gaps, encouraging skill and competition. A leaderboard ranks players, enhancing engagement, with detailed scoring mechanics elaborated in Chapter 6. This approach prioritizes accessibility and feedback, aligning with gamification principles [7].

3.3 Data Collection and Image Selection

The game’s data strategy centers on the DEW dataset, which comprises 1,029,710 Flickr images, from which 10,150 are selected to span the period 1930–1999. The dataset’s metadata (stored in `meta.csv`) provides essential fields—`img_id`, `GT`, `date_taken`, `date_granularity`, and `url`—enabling precise year assignment. Images are retrieved using the DEW-Downloader tool [12], prioritizing diversity in time, geography, and content. The selection process involves sampling to balance variety and difficulty, guided by `GT` and `date_granularity`, with specifics of curation and pre-processing reserved for Chapter 4.

While the DEW dataset contains 155,971 images published under open-access licenses, the image selection process did not explicitly filter based on the `license` or `license_url` fields. Consequently, the 10,150 images used may include items with restrictive or unclear usage rights. Future iterations should incorporate automated license-aware filtering to ensure compliance with open-access criteria and proper at-

tribution requirements. This approach will ensure a robust, historically authentic image pool tailored to the game’s objectives while addressing important intellectual property considerations.

3.4 Challenges in Age Estimation

Several challenges underpin the age estimation task, shaping the game’s design. The DEW dataset’s variable `date_granularity` (ranging from 0–8) introduces inconsistencies, requiring standardized year assignments and scoring flexibility. Visual ambiguity—due to faded images or a lack of temporal cues—complicates both user guesses and system calibration. Human perception varies widely, influenced by subjective cues, which impacts guess accuracy across modes. Additionally, the 10,150-image subset limits diversity, and URL-based retrieval risks access issues. These challenges inform mitigation strategies—such as metadata validation, difficulty tiering, and dataset maintenance—detailed in later chapters, ensuring a balanced and enjoyable experience.

4 Methodology

This chapter details the technical methodology employed in developing "Guess the Age of Photos". Building on the conceptual framework outlined in Chapter 3, this section describes the development environment, implementation strategies, user interface design, and data management processes. We used Python with Flask, Bootstrap, and PostgreSQL to deliver a robust and user-friendly experience.

4.1 Development Environment and Technologies Used

The development of "Guess the Age of Photos" was conducted in a structured environment optimized for web application creation and data handling. The primary programming language was Python, selected for its versatility, extensive libraries, and compatibility with the DEW-Downloader tool used to retrieve dataset images (TIB-Visual-Analytics, 2025 [12]). Flask, a lightweight Python web framework, was employed for its simplicity and flexibility in building RESTful applications, facilitating rapid prototyping and deployment of the game's server-side logic.

For the frontend, Bootstrap was chosen as the CSS framework to ensure responsive, mobile-friendly design with pre-built components, reducing development time while maintaining aesthetic consistency. PostgreSQL (version 15.8) served as the relational database management system (RDBMS), valued for its robustness, support for structured data, and efficient querying capabilities, critical for managing

the game’s image metadata and user information. Development occurred on a local machine running a standard operating system (Windows), with dependencies managed via pip and a virtual environment to ensure reproducibility. Version control was handled with GitLab, hosted on a repository named Bachelor Project in GitLab [8], enabling collaborative tracking of the project’s evolution.

4.2 Frontend and Backend Implementation

The implementation of *Guess the Age of Photos* follows a client-server architecture, ensuring seamless integration between the frontend and backend components.

The **backend**, developed using Flask, manages essential functionalities such as user authentication, game logic, and data retrieval. Key Flask routes include:

- `/login` and `/register`: Handle user authentication, collecting name, age, and gender.
- `/demo`: Allows unregistered users to play without creating an account.
- Various endpoints for mode selection (“Guess the Year” or “Timeline Challenge”), score submission, leaderboard retrieval, and administrative tasks.

The game logic, implemented in Python, processes user inputs, evaluates guesses against the **ground truth (GT) year** from the DEW dataset, and calculates scores accordingly. Flask’s session management ensures smooth user progress tracking within a browser session.

The **frontend**, built with Bootstrap, dynamically renders HTML templates using Flask’s Jinja2 engine. JavaScript enhances interactivity by handling real-time input validation (e.g., ensuring year guesses fall within **1930–1999**) and facilitating AJAX-based requests to update images and scores without requiring a full page reload.

To optimize performance, PostgreSQL is used for structured data management,

interacting with Flask via the `psycpg2` library. The application is deployed on a university server, using Gunicorn as the WSGI server to efficiently handle concurrent requests, ensuring scalability for multiple users.

4.3 User Interface and Experience Design

The design of the user interface (UI) and user experience (UX) prioritizes simplicity, responsiveness, and engagement, aligning with the game’s interactive and educational goals. Upon accessing the platform, users are presented with a Bootstrap-styled landing page featuring a “Demo Mode” button alongside registration and login options.

After logging in, the main interface offers two game modes: “Guess the Year” and “Timeline Challenge”. These are presented in a card-based layout for visual clarity, ensuring ease of selection. Bootstrap’s grid system ensures that the UI remains adaptive across different screen sizes, from desktops to mobile devices.

Guess the Year Mode:

- Displays a single DEW image at the center of the screen.
- Includes a numeric input field (restricted to 1930–1999) for the user’s guess.
- A “Submit” button allows the user to confirm their answer.
- After submission, a Bootstrap modal displays the GT year, actual year, and awarded score, using alert components for immediate feedback.

Timeline Challenge Mode:

- Presents two images side by side using Bootstrap’s row/column structure.
- Users select “Left is Older” or “Right is Older” to indicate which image they believe was taken earlier.
- A feedback modal reveals the correct order and the user’s score.

4 Methodology

A dedicated “Leaderboard” page, accessible via the navigation bar, displays the top scores in a responsive table. To maintain user privacy, only usernames and total scores are shown, without detailed performance statistics.

User Experience Enhancements:

- Minimal clicks required to start playing, ensuring an intuitive flow.
- Tooltips provide concise explanations of game modes.
- Bootstrap’s color scheme creates a visually engaging and dynamic aesthetic.
- Quick AJAX responses ensure smooth interactivity without overwhelming users.
- Accessibility features, such as keyboard navigation and alternative text for images, are implemented to improve inclusivity.

This UI/UX design ensures a visually appealing and user-friendly experience while maintaining efficient gameplay mechanics.

4.4 Data Infrastructure and Image Processing

The game’s data infrastructure is built on PostgreSQL to efficiently manage the DEW dataset’s curated subset and handle user interactions. Given the dataset’s original scale of over one million images, only 155,971 open-access images are available. From this selection, a balanced subset of 10,150 images was curated, ensuring even distribution across the years 1930–1999, with exactly 145 images per year.

4.4.1 Database Design

A structured PostgreSQL schema is designed to support efficient querying and robust data management for the application. The schema comprises three core tables, detailed as follows:

4 Methodology

- **users**: Stores player profiles and administrative details, with the following columns:
 - **id** (Integer, Primary Key): Unique identifier for each user.
 - **name** (Varchar(150), Unique): Player’s unique name.
 - **password** (Varchar(150)): Hashed password for authentication.
 - **gender** (Varchar(150)): User’s gender.
 - **age** (Varchar(150)): User’s age.
 - **security_question** (Varchar(150)): Security question for account recovery.
 - **security_answer** (Varchar(150)): Answer to the security question.
 - **is_admin** (Boolean, Default: False): Indicates if the user has administrative privileges.
 - **date_created** (Timestamp with Timezone, Default: NOW()): Account creation timestamp.
 - **points** (Integer, Default: 0): Cumulative points earned across all games.
 - **dynamic_points** (Double Precision, Default: 0): High-precision score for dynamic calculations.
- **images**: Contains metadata for the image dataset, supporting game functionality, with columns:
 - **id** (BigInteger, Primary Key): Unique identifier for each image.
 - **img_id** (Varchar(255), Unique): Unique identifier from the image source (e.g., Flickr).
 - **gt_year** (Integer): Ground truth year of image acquisition.
 - **date_taken** (Timestamp): Date and time the image was captured.

4 Methodology

- `date_granularity` (Integer): Precision level of the `date_taken` field.
 - `url` (Varchar(255)): URL of the original image.
 - `username` (Varchar(255)): Name of the user who uploaded the image.
 - `title` (Varchar(255)): Descriptive title of the image.
 - `license` (Varchar(255)): Licensing information for the image.
 - `license_url` (Varchar(255)): URL to the license details.
 - `downloaded` (Boolean, Default: False): Indicates if the image has been locally cached.
 - `times_shown` (Integer, Default: 0): Number of times the image has been displayed in games.
 - `times_guessed_correctly` (Integer, Default: 0): Number of correct guesses for the image.
- **game_plays**: Logs player interactions and game outcomes, with the following fields:
 - `id` (Integer, Primary Key): Unique identifier for each game play record.
 - `user_id` (Integer, Foreign Key: `users.id`, On Delete: CASCADE): References the user who played.
 - `image_id` (BigInteger, Foreign Key: `images.id`): References the first image in the game.
 - `second_image_id` (BigInteger, Foreign Key: `images.id`, Nullable): References the second image, if applicable.
 - `game_type` (Varchar(50)): Type of game played (e.g., "Guess the Year" or "Order Images").
 - `is_correct` (Boolean): Indicates if the player's response was correct.

4 Methodology

- `points_earned` (Integer): Points awarded for the game play.
- `correct_order` (Varchar(50), Nullable): Correct sequence of images (if applicable).
- `user_order` (Varchar(50), Nullable): Player’s submitted sequence (if applicable).
- `timestamp` (Timestamp with Timezone, Default: NOW()): Time of the game play.
- `dynamic_score` (Float, Default: 0.0): Additional scoring metric for dynamic evaluation.

Relationships between tables are enforced via foreign keys, ensuring referential integrity. The `game_plays` table links `users` and `images`, supporting one-to-many relationships where a user can have multiple game plays, and an image can appear in multiple game plays as either the first or second image.

4.4.2 Image Selection and Storage

Images were retrieved using the DEW-Downloader tool:

```
python3 downloader.py -i meta.csv -o ~/finalimages/
```

A filtering script ensured that only open-access images from 1930–1999 were selected, distributing 145 images per year for even coverage. Instead of storing images in the project directory, they are placed in `/home/user/finalimages/` on the server. This structure optimizes retrieval speed, reducing latency during gameplay.

4.4.3 Image Processing and Game Integration

To improve performance, all images are preprocessed using Python’s Pillow library:

- Resized to 800x600 pixels for faster loading.

4 Methodology

- Stored efficiently on the server with filenames indexed by `img_id`.
- Retained original URLs in the database for redundancy.

During gameplay, PostgreSQL queries dynamically fetch images:

```
SELECT * FROM images ORDER BY RANDOM() LIMIT 1;
```

For "Timeline Challenge," two images are retrieved:

```
SELECT * FROM images ORDER BY RANDOM() LIMIT 2;
```

A scheduled script periodically verifies URL availability and updates the database to handle broken links. This robust infrastructure ensures smooth gameplay while maintaining historical accuracy and data integrity.

5 Server Deployment and Infrastructure

This section outlines the setup and deployment process for hosting the *Guess the Age of Photos* platform, ensuring it runs smoothly and remains accessible to users.

5.1 Server Environment

The platform is hosted on a university server within the university network, accessible via the domain `disc-imageguessing.uibk.ac.at` or the IP address `138.232.18.56`.

The server has the following specifications:

- **Operating System:** Debian 12
- **CPU:** 4 cores
- **Memory:** 32 GB
- **Storage:** 20 GB

These resources support multiple users at once and provide a reliable experience, even during busy periods. Root access is available on this server via SSH for setup and maintenance.

5.2 Software Setup

The application uses a straightforward software setup:

- **Python and Flask:** For running the web application.
- **PostgreSQL:** For managing the database.
- **Gunicorn:** To handle user requests efficiently.
- **Nginx:** To serve the site and manage traffic.
- **Additional Tools:** Libraries like Pillow for image processing and Flask-Login for user authentication.

These components were installed manually on the server to keep the system lightweight and tailored to the project's needs.

5.3 Database Configuration

PostgreSQL was chosen for its reliability in storing game data, such as user scores and image details. The database is named `thesis`. The schema was designed to quickly retrieve leaderboard rankings and user gameplay history, ensuring smooth performance during play.

5.4 Application Deployment

The deployment process was kept simple:

1. The project files were copied from a local machine to the server (located at `/path/to/Bachelor_Project`), rather than pulled directly from GitLab.
2. A virtual environment was created to manage Python dependencies.
3. Required software was installed from a `requirements.txt` file.

5 Server Deployment and Infrastructure

4. A folder for images (`/home/user/finalimages/`) was created and permissions were set for the web server to access it.

For backups, the project was stored on GitLab, but updates were manually transferred to the server instead of using GitLab's pull functionality.

5.5 Running the Application

Gunicorn serves the application with 4 workers, matching the server's 4-core CPU for balanced performance. It runs as a background service, started with:

```
/path/to/Bachelor_Project/venv/bin/gunicorn --workers 4 --bind  
unix:photo_age_game.sock wsgi:app
```

Nginx acts as a reverse proxy, directing traffic to Gunicorn and serving static files (e.g., images) directly. Its configuration points to the domain `disc-imageguessing.uibk.ac.at` and handles requests efficiently.

5.6 Service Activation

Services were activated with basic commands:

- Start PostgreSQL and Gunicorn.
- Enable them to run on server boot.
- Restart Nginx after linking its configuration.

5.7 Updates and Maintenance

To update the platform:

1. Copy updated files from the local machine to the server via SSH.

5 Server Deployment and Infrastructure

2. Activate the virtual environment.
3. Install any new dependencies.
4. Restart Gunicorn with `sudo systemctl restart`.

For maintenance:

- **Logs:** Nginx and Gunicorn logs are checked regularly.
- **Backups:** Weekly database backups are created using `pg_dump`.

This setup ensures *Guess the Age of Photos* remains accessible, stable, and easy to maintain within the university network.


6 Game Modes

This chapter delineates the operational design and mechanics of "Guess the Age of Photos", our interactive web-based platform that challenges users to estimate the age of historical photographs. The game features two distinct modes, each offering a unique approach to engaging with historical imagery:

6.1 Guess the Age of a Single Photograph

The *Guess the Year* mode invites players to estimate the acquisition year of a single historical photograph, challenging them to interpret visual cues such as clothing styles, technological artifacts, and environmental contexts within the 1930–1999 timeframe. This mode serves as a foundational exercise in temporal deduction, drawing on the DEW dataset's rich metadata to provide an authentic historical experience.

Guess the Year



What year was this photo taken?

Submit Guess

Static Score: 1203

Dynamic Score: 649.3943478260868

Figure 6.1: An example of Guess the Year mode

6.1.1 Gameplay Mechanics

Accessible via the `/guess_the_year` route, the Flask backend renders the `guess_the_year.html` template for authenticated users, ensuring a clean session state by removing demo-related variables:

```
session.pop('demo_mode')
session.pop('demo_tries')
```

An AJAX call to `/get_random_image` retrieves a random `Image` object from the PostgreSQL database using:

```
Image.query.order_by(func.random()).first() (6.1)
```

6 Game Modes

This query selects from the 10,150-image subset, with the image's `img_id` (e.g., 0123456789) mapped to a file path (e.g., 0/12/0123456789.jpg). The Bootstrap frontend displays the image alongside a numeric input field, constrained via HTML5 attributes `min=1930` and `max=1999`. JavaScript enhances interactivity, validating inputs client-side and triggering submissions without page reloads.

Players submit their guess via a POST request to `/check_guess`, sending JSON data (`guess`, `image_id`). The backend retrieves the image's `gt_year`, computes the absolute difference, and logs the attempt in the `GamePlay` table with fields `user_id`, `image_id`, `is_correct`, `points_earned`, and `dynamic_score`.

6.1.2 Scoring System

The scoring system combines static and dynamic points to reward accuracy and proximity. Static points are binary: 10 if the guess is within ± 5 years of `gt_year`, otherwise 0. Dynamic points are calculated using:

$$\text{Dynamic Points} = 10 \times \left(1 - \frac{\min(|\text{guess} - \text{actual_year}|, 100)}{100} \right) \quad (6.2)$$

For example:

- Guess: 1955, Actual: 1950 $\Rightarrow |5| \leq 5$, Static: 10, Dynamic: $10 \times (1 - 5/100) = 9.5$
- Guess: 1960, Actual: 1950 $\Rightarrow |10| > 5$, Static: 0, Dynamic: $10 \times (1 - 10/100) = 9.0$
- Guess: 1980, Actual: 1950 $\Rightarrow |30|$, Static: 0, Dynamic: $10 \times (1 - 30/100) = 7.0$

The ± 5 -year threshold balances leniency with precision, while the dynamic formula ensures partial credit for near-misses. Correct guesses increment `current_user.points` and `dynamic_points` in PostgreSQL.

6.1.3 Feedback and Learning

Post-submission, a JSON response delivers immediate feedback, rendered in a Bootstrap modal with:

- Actual year (`gt_year`)
- Static and dynamic points earned
- Image metadata (`title` or “Untitled”)
- Updated user totals (`total_static_points`, `total_dynamic_points`)

This feedback loop allows players to correlate visual cues with historical contexts, refining their estimation skills.

6.1.4 Demo Mode

For unregistered users, `/guess_the_year_demo` offers a trial experience, rendering the same template with `demo=True`. The `/check_guess_demo` endpoint mirrors full gameplay but tracks session variables instead of database updates, with a five-attempt limit. After five tries, users are prompted to register.

6.2 Compare Two Photographs: Which is Older?

The “Timeline Challenge” mode presents two photographs side-by-side, tasking players with identifying the older image. This comparative approach fosters a nuanced understanding of chronological progression, leveraging the DEW dataset’s temporal diversity to create varied and thought-provoking pairs.

6 Game Modes

Timeline Challenge

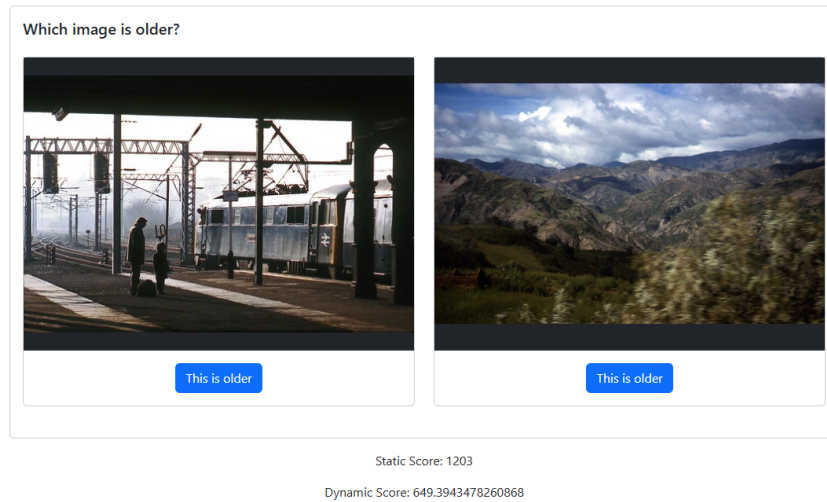


Figure 6.2: An example of the Timeline Challenge mode

6.2.1 Gameplay Mechanics

Initiated at `/timeline_challenge`, the Flask backend renders `timeline_challenge.html`, clearing demo session variables for logged-in users. The `/get_random_image_pair` endpoint fetches two images via:

```
Image.query.filter(Image.gt_year.isnot(None)).all()
```

A randomization loop (up to 5 attempts) ensures distinct `img_id` and `gt_year` values, constructing paths (e.g., `0/12/0123456789.jpg`) served from `/static/images`. The Bootstrap frontend arranges images in a row with `col-md-6` columns, displaying “Left is Older” and “Right is Older” buttons (`btn-outline-secondary`) below each pair. JavaScript disables the opposing button post-selection, preventing duplicate submissions, and submits via AJAX to `/check_timeline_guess`.

The POST request sends JSON (`selected_id`, `other_id`, `selected_newer`), logging results in `GamePlay` with `second_image_id` for pair tracking. This design ensures a fluid, visually balanced experience, optimized for comparative analysis.

6.2.2 Scoring System

Correctness hinges on the `selected_newer` flag:

- `selected_newer=False`: Correct if `selected_year < other_year`
- `selected_newer=True`: Correct if `selected_year > other_year`

Static points award 10 for correct answers, 0 otherwise.

To better reflect the difficulty of each comparison, the system introduces dynamic scoring via `calculate_dynamic_points(year_diff)`, where $\Delta = |\text{selected_year} - \text{other_year}|$. This metric quantifies the temporal gap between two images, under the assumption that distinguishing between two closely dated images is substantially harder than when the years are far apart.

$$\text{Dynamic Points} = 5.0 \times \begin{cases} 1.0 & \text{if } \Delta \leq 10 \\ 1.0 - \frac{\Delta-10}{40} & \text{if } 10 < \Delta \leq 50 \\ 0.1 & \text{if } \Delta > 50 \end{cases}$$

This tiered function reflects the idea that:

- When two images are taken within a short time span ($\Delta \leq 10$), distinguishing the newer one is challenging and deserves full dynamic reward.
- As the gap increases ($10 < \Delta \leq 50$), the task becomes easier, and dynamic points linearly decrease to discourage rewarding trivial comparisons.
- For very large gaps ($\Delta > 50$), only a minimal reward is given, since such comparisons are considered obvious and do not reflect high player skill.

Examples include:

- $\Delta = 5$: $5.0 \times 1.0 = 5.0$ (Static: 10 if correct)
- $\Delta = 25$: $5.0 \times (1 - \frac{25-10}{40}) = 5.0 \times 0.625 = 3.12$ (Static: 10 if correct)

6 Game Modes

- $\Delta = 60$: $5.0 \times 0.1 = 0.5$ (Static: 10 if correct)

This scoring system rewards not only correctness but also difficulty, creating an incentive to perform well even on harder rounds and adding a nuanced layer to the game’s competitive design. Correct guesses update both `current_user.points` and `dynamic_points`, which are then committed to the database for tracking user performance.

6.2.3 Feedback and Learning

Feedback, delivered via JSON and displayed in a Bootstrap modal (`alert-success` for correct, `alert-danger` for incorrect), includes:

- Both images’ `gt_year` values
- Static and dynamic points
- Metadata for both images (title)

This detailed response enables players to analyze chronological shifts—e.g., a 1940s black-and-white image versus a 1970s color photo—enhancing their ability to discern subtle temporal markers. The side-by-side presentation and immediate comparison foster relational learning, complementing the individual focus of “Guess the Year.”

6.2.4 Demo Mode

The `/timeline_demo` route provides a demo experience, setting `session['demo_mode'] = True` and `session['demo_tries'] = 500`. The `/check_timeline_guess_demo` endpoint processes guesses similarly, updating session scores without database writes. Limited to five attempts, it mirrors the full mode’s mechanics, using the same DEW image pool to maintain authenticity while enticing users to register.

6.3 Ranking and Scoring Mechanism

The ranking and scoring mechanism integrates both modes into a cohesive competitive framework, implemented across gameplay endpoints and displayed via `/scoreboard`. This system leverages detailed logging and dual-point tracking to enhance motivation and provide actionable insights.

6.3.1 Point Tracking

Two point types are maintained:

- Static Points (`User.points`): Binary (10 or 0), reflecting correctness, updated in `check_guess` and `check_timeline_guess`.
- Dynamic Points (`User.dynamic_points`): Scaled by accuracy or difficulty, offering granular performance feedback.

Each play is recorded in the `GamePlay` table:

```
(user_id, image_id, game_type, is_correct, points_earned,  
dynamic_score, second_image_id)
```

For “Timeline Challenge,” `second_image_id` and `correct_order/user_order` fields capture pair-specific data, enabling comprehensive analytics. This structure supports both real-time scoring and post-game analysis, stored efficiently in PostgreSQL.

6.3.2 Leaderboard

The `/scoreboard` route generates dual leaderboards:

- Static: `User.query.order_by(User.points.desc())`
- Dynamic: `User.query.order_by(User.dynamic_points.desc())`

6 Game Modes

Rankings handle ties by preserving the same rank for equal scores (e.g., two users at 50 points both rank 1), computed via previous-point comparisons. Rendered in `scoreboard.html` with Bootstrap table components, entries display anonymized names (e.g., “HasanY”), balancing transparency with privacy. The dual-ranking approach highlights different skills—consistency (static) versus precision (dynamic)—enhancing competitive depth.

6.3.3 Gamification Elements

The system incorporates gamification principles [7]:

- Immediate Feedback: Modals post-guess provide instant results, reinforcing learning.
- Progressive Difficulty: Dynamic scoring rewards tighter guesses or smaller Δ , scaling challenge with skill.
- Social Comparison: Leaderboards foster rivalry, updated in real-time via Unicorn’s concurrency on the university server.
- Achievement Tracking: Separate static and dynamic totals offer tangible progress markers.

Additional analytics, accessible via `/admin/dashboard`, filter performance by age and gender, displaying top images and aggregate stats (e.g., `guess_year_total_correct`). This rich feedback ecosystem encourages repeated play, skill refinement, and historical curiosity, leveraging the DEW dataset’s authenticity to create a compelling experience.

6.3.4 Points Aggregation Across Game Modes

Points are computed cumulatively across both game modes—*Guess the Year* and *Timeline Challenge*—and stored in shared fields such as `points` and `dynamic_points`.

6 Game Modes

This design ensures a unified scoring experience that rewards overall engagement and accuracy. While point tracking is consolidated, the dashboard enables disaggregated views, allowing users to see total points as well as mode-specific contributions (e.g., `guess_the_year_points` vs. `timeline_challenge_points`). This promotes self-reflection, encourages diversified gameplay, and maintains fairness across different playing styles.

Two visualizations in the dashboard support this transparency: Figure 6.3 shows the live scoreboard with both ranking types, while Figure 6.4 presents the user’s point distribution chart broken down by game mode. These visuals enhance usability and foster healthy competition by making performance immediately visible.

Scoreboard

Static Scoring				Dynamic Scoring			
Rank	Username	Static Points	Performance	Rank	Username	Dynamic Points	Performance
1	Ali	5150	View Performance	1	Joachim	1601.06	View Performance
2	Joachim	5000	View Performance	2	Lisa Marie	1551.43	View Performance
3	Lisa Marie	4780	View Performance	3	Ali	1480.16	View Performance
4	Ben	4440	View Performance	4	Ali	1381.32	View Performance
5	Applepie	4250	View Performance	5	Olivia	1276.86	View Performance
6	Ali	4050	View Performance	6	Applepie	1176.25	View Performance
7	Freddie	3740	View Performance	7	Freddie	1120.20	View Performance
8	Sara	3700	View Performance	8	Sara	1076.99	View Performance
9	Leonie	3300	View Performance	9	Martin Mueller	1005.60	View Performance
10	Mads	3000	View Performance	10	Mads	1000.03	View Performance
11	Oscar	2700	View Performance	11	Oscar	1011.54	View Performance
12	Kianle	2620	View Performance	12	Leonie	986.04	View Performance
13	Hans Willem	2500	View Performance	13	Kianle	886.21	View Performance
14	alaska	2450	View Performance	14	alaska	791.79	View Performance
15	Martin Mueller	2380	View Performance	15	Hans Willem	786.13	View Performance

Figure 6.3: Live scoreboard with static and dynamic point rankings.

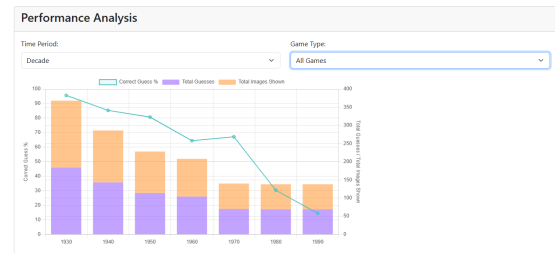


Figure 6.4: User’s point breakdown by game mode.

7 User Instructions and Cohort Management

To provide a comprehensive understanding of the user experience and evaluation methodology, this section details the instructions provided to users and how the participant cohort was managed throughout the study.

7.1 User Instructions

Users were provided with comprehensive information about the platform through multiple accessible channels:

7.1.1 Our Mission Tab

All users had access to the "Our Mission" tab on the platform. This section provided detailed information about:

- The specific datasets used in the study
- The research objectives and purpose of the platform
- The broader context and significance of the project

This transparency ensured users understood the scientific foundation of the platform and how their participation contributed to research goals.

7.1.2 How to Play and Point System

On the landing page, users found a prominently displayed button labeled "How to Play and Point System." When clicked, this button revealed:

- Comprehensive rules for each game mode
- Detailed explanation of the scoring system
- Tips for improving performance
- Information about leaderboards and achievements

This user-friendly approach ensured that gameplay instructions were easily accessible while keeping the main interface clean and focused.

7.2 Cohort Management

Our participant cohort was assembled through targeted recruitment across several key demographics:

Recruitment Methods

The participant cohort was assembled through multiple channels to ensure demographic diversity:

- **Primary Recruitment:** In-person visits to the researcher's former high school, involving direct classroom presentations and teacher coordination (60% of participants)
- **Secondary Recruitment:** University computer science department channels via Whatsapp targeting students across multiple academic years (2022/23, 2023/24, and 2024/25) (30% of participants)
- **Tertiary Recruitment:** Personal networks including friends, family members, and extended connections of the researcher (10% of participants)

7.2.1 Participant Demographics

The final cohort consisted of 113 registered users with the following characteristics:

- **Age Distribution:**

- 14-18 years: 67 users (59.3%)
- 19-24 years: 29 users (25.7%)
- 25-30 years: 7 users (6.2%)
- 31-40 years: 3 users (2.7%)
- 41+ years: 7 users (6.2%)

- **Gender Distribution:**

- Male: 64 users (56.6%)
- Female: 49 users (43.4%)

- **Educational Background:**

- Secondary education: 67 users (59.3%)
- Undergraduate (Computer Science): 32 users (28.3%)
- Other backgrounds: 14 users (12.4%)

7.2.2 Participation Guidelines

Users were provided with the following guidelines upon registration:

- No minimum participation requirement was enforced, allowing for natural engagement patterns
- Users were encouraged to try both game modes at least once
- Feedback survey participation was optional but encouraged after 10+ gameplay sessions

7 User Instructions and Cohort Management

- Users were informed that their anonymized data would be analyzed for research purposes

8 Evaluation

This chapter presents a comprehensive evaluation of *Guess the Age of Photos*, drawing on data collected as of March 30, 2025, at 13:45. The analysis explores four key areas: the testing setup, user feedback and experience, accuracy of user guesses, and engagement and retention metrics. With 113 registered users contributing to 15,473 gameplays across two modes—*Guess the Year* and *Timeline Challenge*—this assessment showcases the platform’s ability to deliver an entertaining yet educational experience. By examining these dimensions, we gain insights into its usability, appeal, and potential for future enhancements.

8.1 Testing Setup

The evaluation was conducted on the live version of *Guess the Age of Photos*, deployed on a university server powered by Gunicorn, a Python web server, and PostgreSQL (version 15.8) for robust data storage and retrieval. The testing period spanned from the platform’s launch, 20.01.2025, to March 30, 2025, capturing real-world interactions from users who discovered it organically—no controlled testing groups were involved. The platform leveraged a dataset of 10,150 images sourced from the Date Estimation in the Wild (DEW) dataset, covering the years 1930 to 1999. Each image’s ground truth year (`gt_year`) was meticulously validated using metadata, ensuring reliable benchmarks for user guesses.

Users accessed the platform through standard web browsers on a variety of devices,

8 Evaluation

including desktops, laptops, and mobile phones. This broad compatibility was enabled by Bootstrap, a front-end framework that ensures a responsive, user-friendly design across screen sizes. Every gameplay action was logged in the **GamePlay** table, recording essential details such as `user_id`, `image_id`, guess accuracy, and scores. Additionally, a feedback button on the home tab invited users to rate their experience via five questions, scored on a 1-to-5 scale (1 = very good, 5 = very bad). These responses provided qualitative insights to complement the quantitative data, painting a fuller picture of the platform’s performance.

8.2 User Feedback and Experience

To understand how users perceived *Guess the Age of Photos*, feedback was collected from 20 to 24 respondents via the home tab survey. This section analyzes their ratings across five categories, providing insights into gameplay balance, enjoyment, and usability.

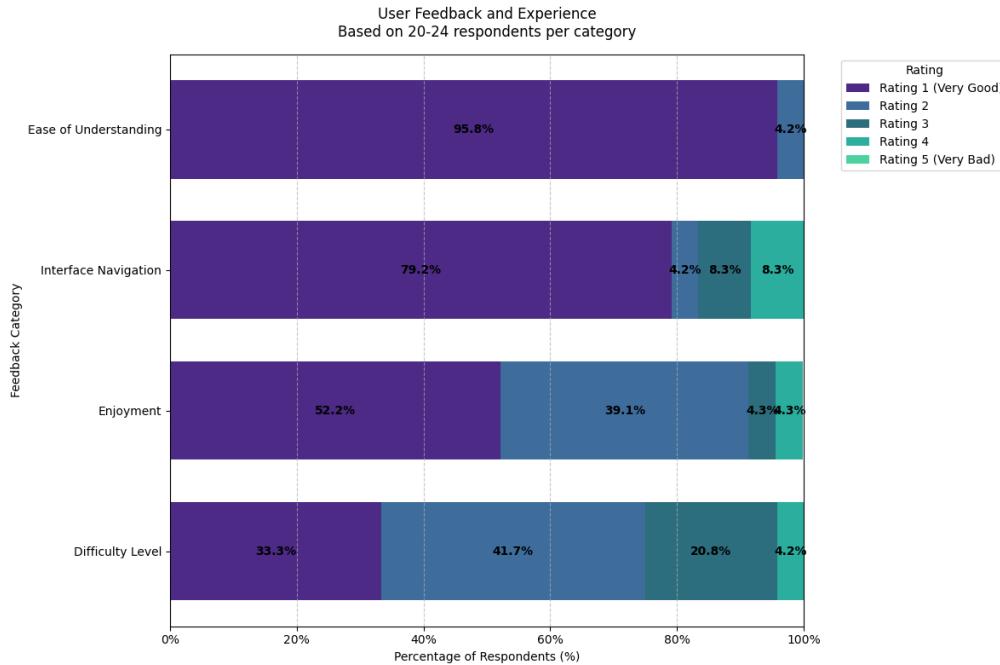


Figure 8.1: User feedback ratings across five evaluation categories

8 Evaluation

As shown in Figure 8.1, user feedback was predominantly positive across all evaluation categories:

- **Difficulty Level Balance:** Out of 24 responses, 33.3% (8 users) gave a rating of “1” (very well-balanced), 41.7% (10 users) a “2”, 20.8% (5 users) a “3”, and 4.2% (1 user) a “4”. With 75% of users selecting “1” or “2”, the difficulty appears well-calibrated for most. However, the 25% who rated it “3” or “4” suggest that some images might pose unexpected challenges.
- **Enjoyment:** From 23 responses, 52.2% (12 users) rated enjoyment as “1” (highly enjoyable), 39.1% (9 users) as “2”, and 4.3% (1 user each) as “3” and “4”. This overwhelming 91.3% positive response (“1” or “2”) underscores the platform’s appeal. Features like leaderboards, which foster competition, and immediate feedback after guesses likely amplify this enjoyment.
- **Interface Navigation:** Among 24 respondents, 79.2% (19 users) rated navigation as “1” (very easy), 4.2% (1 user) as “2”, and 8.3% (2 users each) as “3” and “4”. The strong approval—over three-quarters at “1”—highlights the success of the intuitive layout. Bootstrap’s responsive design clearly adapts seamlessly to diverse devices, though the small minority of lower ratings may point to occasional hiccups, such as mobile-specific quirks.
- **Ease of Understanding:** Of 24 responses, an impressive 95.8% (23 users) rated this as “1” (very easy), with just 4.2% (1 user) at “2”. This near-unanimous praise reflects how clearly the game mechanics and instructions are presented—users can jump in with minimal confusion, a critical factor for broad accessibility.
- **Overall Rating:** Across 20 responses, the average rating was 4.25 out of 5. This high score signals robust satisfaction, confirming that the platform effectively blends entertainment with education, as intended.

Overall, the feedback paints a picture of a highly usable and enjoyable platform.

The slight variability in difficulty perception offers a hint for improvement—perhaps adaptive mechanics could tailor challenges to individual skill levels.

8.3 Accuracy of User Guesses

Accuracy provides a window into how well users perform on *Guess the Year* and *Timeline Challenge*. This section analyzes 15,473 gameplays—2,252 for *Guess the Year* and 13,221 for *Timeline Challenge*—with detailed breakdowns by age group and gender.

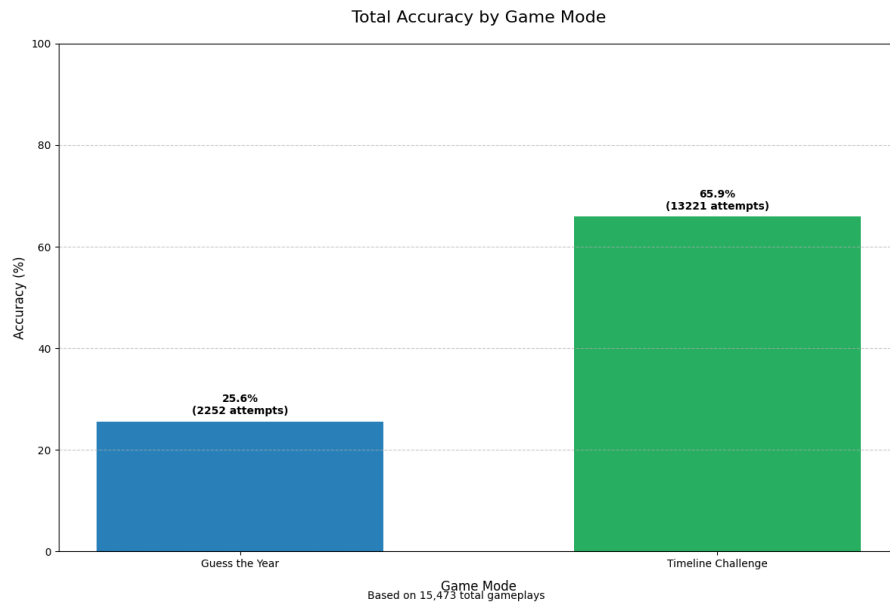


Figure 8.2: Comparison of accuracy rates between game modes

8.3.1 Total Accuracy

As illustrated in Figure 8.2, there is a significant difference in performance between the two game modes:

- *Guess the Year*: Out of 2,252 attempts, 577 guesses were correct, yielding an accuracy of 25.62%. Here, “correct” means the guess fell within ± 5 years of the `gt_year`. This relatively low rate reflects the inherent difficulty of pinpointing

8 Evaluation

an exact year based solely on visual cues—a task requiring precise historical knowledge.

- *Timeline Challenge*: Out of 13,221 attempts, 8,708 sequences were correctly ordered, achieving a 65.9% accuracy rate. This higher success rate suggests that comparing images relatively (e.g., determining which came first) is more intuitive than absolute year estimation, aligning with human cognitive strengths in pattern recognition.

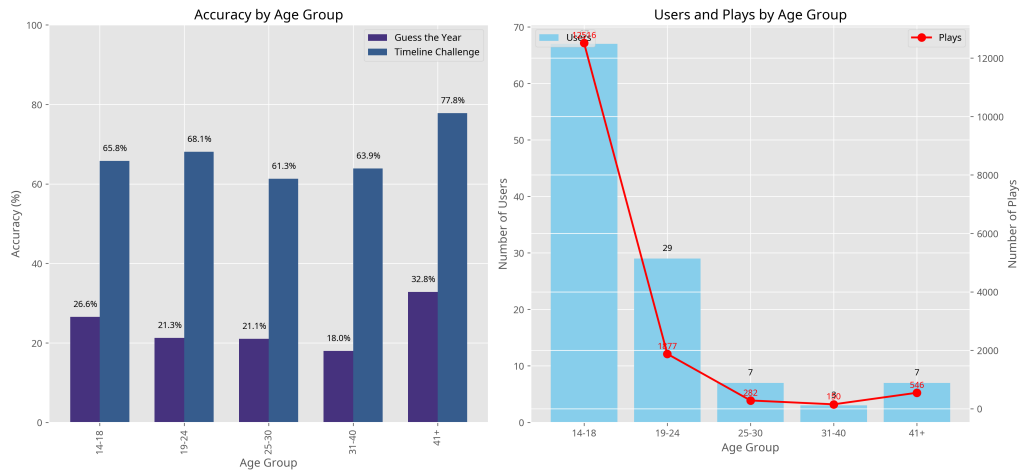


Figure 8.3: Accuracy rates by age group for both game modes

8.3.2 Breakdown by Age Groups

Figure 8.3 visualizes the performance patterns across different age demographics:

- **Ages 14–18** (67 users, 12,516 plays): For *Guess the Year*, accuracy was 26.55% (339 correct out of 1,277), while *Timeline Challenge* scored 65.3% (7,344 out of 11,239). This group’s solid performance may stem from frequent play and familiarity with the platform.
- **Ages 19–24** (29 users, 1,877 plays): *Guess the Year* accuracy dropped to 21.27% (114 out of 546), but *Timeline Challenge* rose to 68.1% (913 out of 1,341). The higher comparative accuracy could indicate stronger analytical

8 Evaluation

skills or exposure to diverse historical contexts.

- **Ages 25–30** (7 users, 282 plays): Accuracy for *Guess the Year* was 21.05% (24 out of 114), and for *Timeline Challenge*, 61.3% (103 out of 168). With fewer users and plays, this group’s smaller sample size limits firm conclusions, but the trend mirrors broader patterns.
- **Ages 31–40** (3 users, 150 plays): *Guess the Year* accuracy was 17.98% (16 out of 89), while *Timeline Challenge* scored 63.9% (39 out of 61). The lowest *Guess the Year* accuracy in this group suggests potential knowledge gaps.
- **Ages 41+** (7 users, 546 plays): *Guess the Year* accuracy was 32.82% (64 out of 195), while *Timeline Challenge* reached 77.8% (273 out of 351).

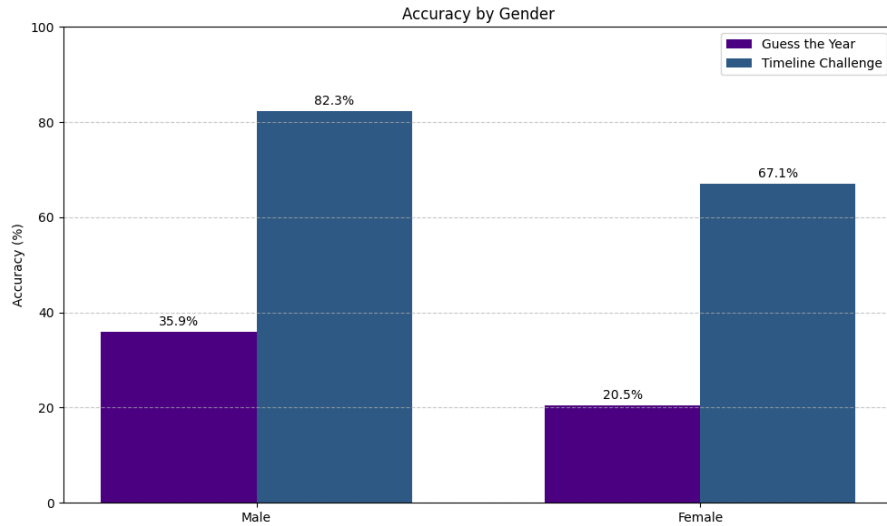


Figure 8.4: Accuracy comparison between male and female participants

8.3.3 Decade-wise Difficulty Across Users

To further assess historical perception, we computed decade-wise accuracy for both game modes across all users. This additional analysis mirrors Figure 8.2, but focuses on identifying which decades users found most challenging. The resulting insights were derived from all recorded attempts, without filtering by age or gender, thus

8 Evaluation

providing a macro-level perspective of difficulty.

The data used for this analysis was exported directly from the PostgreSQL database running on the game server. A CSV containing all gameplay records was generated via SQL and securely transferred to a local machine using the following command:

```
scp hasan@disc-imageguessing.uibk.ac.at:/tmp/gameplays_export.csv .
```

This file included fields such as user ID, game type, whether the guess was correct, and the ground truth year (gt_year) of the image. Using Python and `pandas`, the `gt_year` field was converted into decade buckets by applying floor division: $\text{decade} = \lfloor \frac{\text{gt_year}}{10} \rfloor \times 10$. This groups all years from 1930 to 1939 under the decade “1930,” 1940–1949 under “1940,” and so on. Correct and incorrect counts were then used to calculate accuracy rates for each decade as $\text{accuracy} = \frac{\text{correct}}{\text{total}}$.

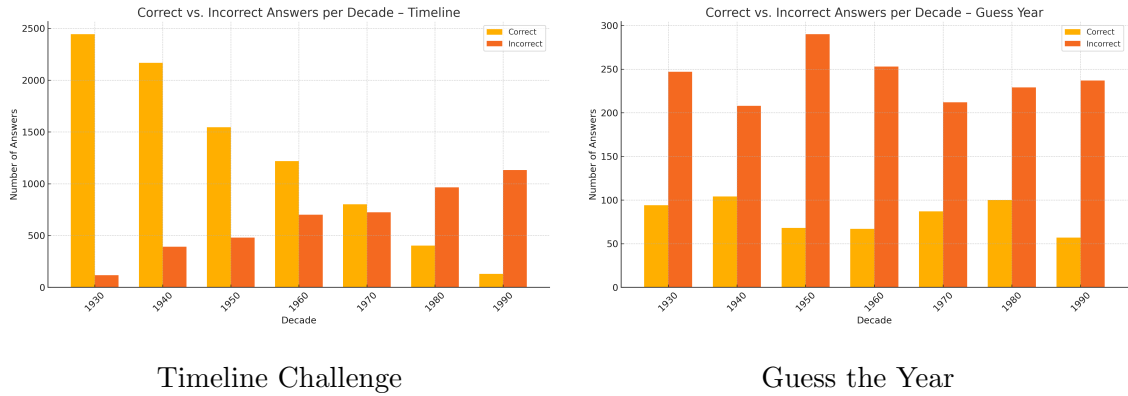


Figure 8.5: Number of correct and incorrect answers per decade by game mode

Timeline Challenge

As shown in Figure 8.5a, users demonstrated high accuracy for older decades, particularly the 1930s and 1940s, where correct guesses dominated. Accuracy gradually decreased with each subsequent decade, with the 1990s proving the most difficult. This trend suggests that users may possess stronger intuition or historical context

for earlier periods, likely due to educational exposure or visual cues more easily linked to classic historical imagery.

Guess the Year

In contrast, Figure 8.5b displays a consistently lower performance across all decades, with correct guesses rarely outnumbering incorrect ones. This is expected, given the challenge of pinpointing exact years. Interestingly, the highest accuracy was achieved for the 1940s and 1980s, though still well below 35%. This supports the earlier claim that relative ordering (Timeline Challenge) is cognitively easier for users than precise dating.

These findings align with cognitive psychology research indicating humans are better at relative comparisons than absolute estimations [13], particularly when dealing with abstract visual information.

8.3.4 Breakdown by Gender

As shown in Figure 8.4, there were notable performance differences between gender groups:

- **Males** achieved accuracy rates of 35.9% in *Guess the Year* and 82.3% in *Timeline Challenge*
- **Females** achieved accuracy rates of 20.51% in *Guess the Year* and 67.1% in *Timeline Challenge*

This performance gap may be attributed to greater familiarity with historical context or differences in prior exposure to historical photography among the participant groups.

These findings highlight the platform’s dual nature: it entertains while subtly testing historical perception. The variation across age groups suggests differing levels of

8 *Evaluation*

historical intuition or engagement, opening the door for future enhancements, such as AI-driven hints to boost accuracy.

8.4 Engagement and Retention Metrics

To gauge how deeply users connect with *Guess the Age of Photos*, this section examines gameplay frequency, user distribution, and repeat play behavior—key indicators of engagement and loyalty.

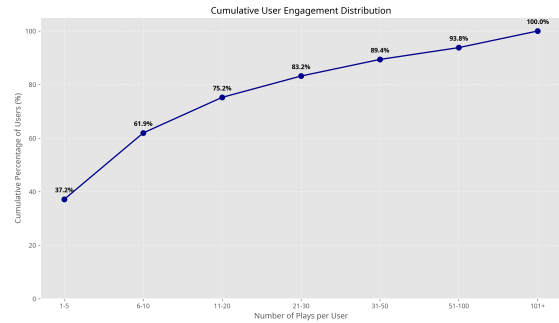
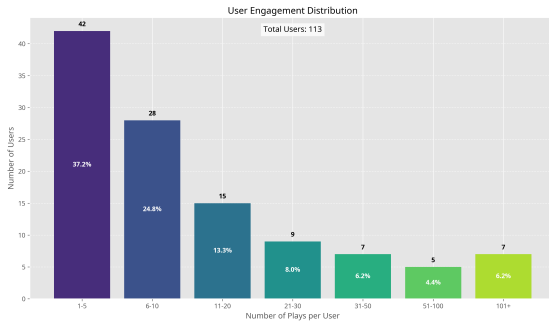


Figure 8.6: Distribution of users by number of gameplay sessions

Figure 8.7: Cumulative percentage of users by gameplay frequency

As visualized in Figure 8.6, user engagement follows a characteristic pattern with a long tail distribution:

- **Casual Users (1-5 plays):** 37.2% (42 users) engaged only briefly with the platform, suggesting either initial curiosity without sustained interest or recent joiners who haven't yet accumulated more plays.
- **Moderate Users (6-20 plays):** 38.1% (43 users) showed meaningful engagement, exploring both game modes and likely experiencing the full range of platform features.
- **Active Users (21-50 plays):** 14.2% (16 users) demonstrated significant commitment to the platform, regularly returning for new gameplay sessions.
- **Power Users (51+ plays):** 10.6% (12 users) exhibited exceptional engagement, with some individuals contributing hundreds of plays. These users form the platform's core community and likely drive competitive aspects through leaderboard participation.

8 Evaluation

Figure 8.7 provides additional insight into the distribution pattern, showing that approximately 75% of users played fewer than 20 times, while the remaining 25% account for the majority of total gameplay activity.

- **Gameplay Distribution:** Of the 15,473 total gameplays, 85.4% (13,221) were for *Timeline Challenge*, dwarfing the 14.6% (2,252) for *Guess the Year*. This lopsided preference likely stems from the comparative mode’s higher success rate and simpler cognitive demand, making it more immediately rewarding.
- **User Activity:** The 14–18 age group dominated, with 67 users accounting for 12,516 gameplays—80.9% of the total. This translates to an average of 186.8 plays per user, a testament to the platform’s appeal among younger audiences.
- **Repeat Play Patterns:** Among the 113 users, the top 10% (11 users) drove 48% of all gameplays (7,424), averaging 674.9 plays each. This concentration reveals a highly dedicated core group, likely motivated by competitive elements like leaderboards, while the broader user base engages more casually.

Together, these metrics illustrate a platform that captivates users, particularly younger ones, and fosters significant repeat engagement among a committed subset. This strong foundation supports its educational and entertainment goals while hinting at opportunities to broaden appeal across age groups and increase retention among casual users.

9 Future Work and Extensions

This chapter outlines potential enhancements for *Guess the Age of Photos*. Building on the existing framework, these proposed extensions aim to enrich gameplay, expand the dataset in a targeted manner, and improve accessibility for a wider audience.

9.1 Key Game Mode and Feature Enhancements

Prioritizing user engagement and providing more varied gameplay experiences, the following game modes and features are considered key areas for near-term development.

9.1.1 Decade Match Mode: Deepening Historical Understanding

The *Decade Match* mode aims to challenge players to think more broadly about historical periods. By presenting several images simultaneously and requiring their assignment to the correct decade (e.g., 1930s, 1950s, 1970s), this mode encourages an understanding of the visual shifts and cultural trends that define different eras. Implementation would involve modifying the backend data handling to manage multiple image-guess pairs within a single game session and creating an intuitive drag-and-drop or selection-based interface on the frontend for decade assignment. The scoring system would need to be adapted to reward accurate decade placements, potentially with bonus points for precise ordering or grouping.

9.1.2 Basic Hint System: Enhancing Accessibility

To make the game more approachable for users who may not have a strong background in historical visual cues, a basic *Hint System* could be introduced. This initial version could provide a single, simple textual hint for each image, such as a prominent object mentioned in the original DEW dataset’s title or description (e.g., "automobile," "fashion," "building"). Players could opt to reveal this hint at the cost of a small reduction in the potential points awarded for a correct guess. Implementing this feature would require adding a hint field to the `Image` database model and modifying the game logic to deduct points when a hint is requested.

9.2 Focused Dataset Expansion: Quality over Quantity

Instead of a massive and potentially overwhelming integration of the entire DEW dataset, a more strategic approach to dataset expansion would focus on quality and thematic relevance.

9.2.1 Curated Themed Collections: Adding Depth and Variety

Introducing smaller, curated collections of images centered around specific themes (e.g., the evolution of aviation, popular hairstyles across the 20th century, major historical events as captured in photographs) would add significant depth and variety to the gameplay. These collections could be sourced from reputable public domain archives, museum digital collections with open access policies, or carefully vetted user submissions. The process would involve identifying relevant themes, sourcing high-quality and well-dated images, ensuring proper metadata attribution, and integrating these collections into the game with appropriate tagging for themed challenges.

9.3 Mobile Accessibility: Reaching a Wider Audience

Recognizing the prevalence of mobile devices, improving the mobile experience is a crucial step towards broader accessibility.

9.3.1 Responsive Web Design Improvements: A Practical First Step

As a more immediate and less resource-intensive approach than developing a full native mobile application, significant improvements to the responsiveness of the existing Flask-based web interface would be undertaken. This would involve refining the layout and styling using Bootstrap's responsive design capabilities to ensure optimal viewing and interaction across different screen sizes and orientations. The focus would be on making the game playable and enjoyable on smartphones and tablets without the need for a dedicated app. This would lay the groundwork for potential future native app development based on user feedback and resource availability.

10 Conclusion

Guess the Age of Photos provides an engaging platform for estimating the historical age of photographs. Utilizing a 10,150-image subset of the DEW dataset, the game’s dual modes—*Guess the Year* and *Timeline Challenge*—implemented with Python, Flask, Bootstrap, and PostgreSQL, promote analytical skills and historical awareness. The dynamic scoring system, balancing correctness with precision-based rewards, combined with a competitive leaderboard deployed on a university server via Unicorn, creates a compelling gamified experience. Immediate feedback and detailed analytics further enhance the learning process.

Beyond its current scope, the game presents significant opportunities for expansion. Future research could compare human intuition against AI precision, analyzing dataset-wide trends and performance metrics. Such a study could illuminate differences in cognitive versus algorithmic perception—e.g., human strengths in cultural context versus AI efficiency in pattern recognition.

Moreover, integrating new game modes, expanding the dataset, and developing a mobile application would further broaden accessibility and engagement. Combining AI advancements, NLP-based contextual reasoning, and reinforcement learning could elevate *Guess the Age of Photos* into a sophisticated educational tool that not only entertains but also contributes to historical scholarship and AI-human interaction studies. These extensions position the project as a foundation for future advancements in both gamification and AI-assisted historical analysis.

Bibliography

- [1] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: Defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference*, pages 9–15. ACM, 2011.
- [2] Adrián Domínguez, José Sáenz-de Navarrete, Luis de Marcos, Luis Fernández-Sanz, Carmen Pagés, and José J. Martínez-Herráiz. Gamifying learning experiences: Practical implications and outcomes. *Computers & Education*, 63:380–392, 2013.
- [3] Guang Guo and Guo Mu. Age estimation from face images: Human vs. machine performance. In *Proceedings of the IEEE International Conference on Biometrics*, pages 1–8. IEEE, 2013.
- [4] Juho Hamari, Jonna Koivisto, and Harri Sarsa. Does gamification work? – a literature review of empirical studies on gamification. In *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS)*, pages 3025–3034. IEEE, 2014.
- [5] He Han, Carl Otto, Xiang Liu, and Anil K. Jain. Demographic estimation from face images: Human vs. machine performance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(6):1148–1161, 2014.
- [6] Eric Müller, Matthias Springstein, and Ralph Ewerth. When was this picture taken? image date estimation in the wild. In *Advances in Information Retrieval: Proceedings of the 39th European Conference on Information Retrieval (ECIR)*,

Bibliography

- volume 10193 of *Lecture Notes in Computer Science (LNCS)*, pages 619–625. Springer, 2017.
- [7] Scott Nicholson. A recipe for meaningful gamification. In Torsten Reiners and Lincoln C. Wood, editors, *Gamification in Education and Business*, pages 1–20. Springer, 2015.
- [8] University of Innsbruck Computer Science Department. Bachelor project repository. <https://git.uibk.ac.at/csba4787/bachelor-project>, 2025. Accessed: 2025-03-11.
- [9] Matthew G. Rhodes. Age estimation of faces: A review. *Applied Cognitive Psychology*, 23(1):1–12, 2009.
- [10] Katie Seaborn and Deborah I. Fels. Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, 74:14–31, 2015.
- [11] Jiwon Shin, Hyeon Kim, and Seungmin Lee. Perceived age estimation from facial image and demographic data in young and middle-aged south korean adults. *Scientific Reports*, 14:78695, 2024.
- [12] TIB Visual Analytics. DEW-Downloader GitHub Repository. <https://github.com/TIB-Visual-Analytics/DEW-Downloader>, 2017.
- [13] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, 1974.