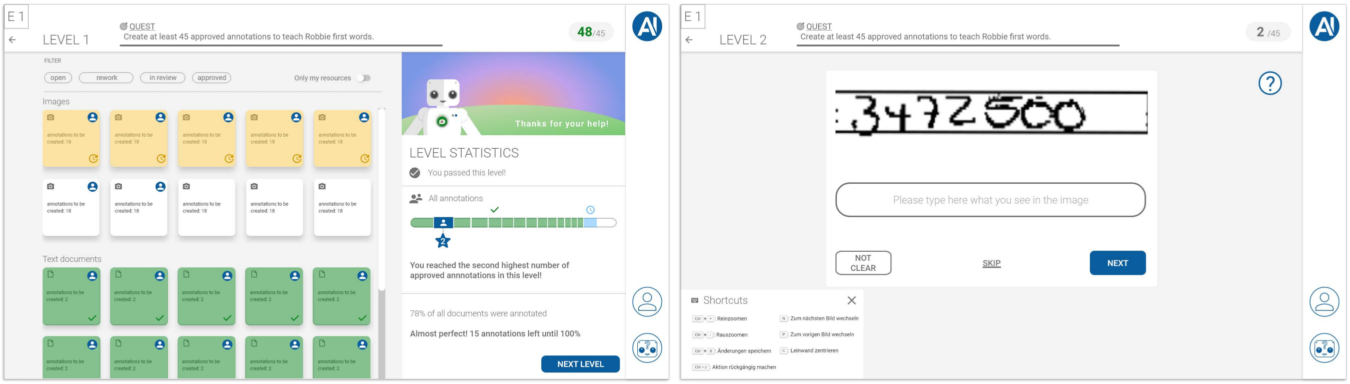
# Achiever or Explorer? Gamifying the Creation Process of Training Data for Machine Learning

Sarah Alaghbari
Annett Mitschick
sarah.alaghbari@tu-dresden.de
annett.mitschick@tu-dresden.de
Technische Universität Dresden
Dresden, Germany

Gregor Blichmann
Martin Voigt
gregor.blichmann@ai4bd.com
martin.voigt@ai4bd.com
AI4BD Deutschland GmbH
Dresden, Germany

Raimund Dachselt
raimund.dachselt@tu-dresden.de
Centre for Tactile Internet with
Human-in-the-Loop (CeTI),
Technische Universität Dresden
Dresden, Germany


E 1
LEVEL 1
QUEST
48/45
FILTER
open
rework
in review
approved
Only my resources
Images
annotations to be created: 18
annotations to be created: 18
annotations to be created: 18
annotations to be created: 18
annotations to be created: 18
annotations to be created: 18
annotations to be created: 18
annotations to be created: 18
annotations to be created: 18
annotations to be created: 18
Text documents
annotations to be created: 2
annotations to be created: 2
annotations to be created: 2
annotations to be created: 2
annotations to be created: 2
annotations to be created: 2
annotations to be created: 2
annotations to be created: 2
annotations to be created: 2
annotations to be created: 2
Thanks for your help!
LEVEL STATISTICS
All annotations
2
NEXT LEVEL
E 1
LEVEL 2
QUEST
2 /45
3472500
NOT CLEAR
SKIP
NEXT
Shortcuts

Create at least 45 approved annotations to teach Robbie first words.

You passed this level!

You reached the second highest number of approved annotations in this level!

78% of all documents were annotated

Almost perfect! 15 annotations left until 100%

Create at least 45 approved annotations to teach Robbie first words.

Please type here what you see in the image

Ctrl + + : Reinzoomen
Ctrl + - : Rauszoomen
Ctrl + S : Änderungen speichern
Ctrl + z : Aktion rückgängig machen
N : Zum nächsten Bild wechseln
P : Zum vorigen Bild wechseln
C : Leinwand zentrieren

**Figure 1: Two screenshots of the gamified annotation tool for the creation of training data for machine learning processes: a passed level (left) and an example annotation task for handwriting recognition (right)**

## ABSTRACT

The development of artificial intelligence, e. g., for Computer Vision, through supervised learning requires the input of large amounts of annotated or labeled data objects as training data. The creation of high-quality training data is usually done manually which can be repetitive and tiring. *Gamification*, the use of game elements in a non-game context, is one method to make tedious tasks more interesting. This paper proposes a multi-step process for *gamifying* the manual creation of training data for machine learning purposes. We choose a user-adapted approach based on the results of a preceding user study with the target group (employees of an AI software development company) which helped us to identify annotation use cases and the users' player characteristics. The resulting concept includes levels of increasing difficulty, tutorials, progress indicators and a narrative built around a robot character which at the same time is a user assistant. The implemented prototype is an extension of the company's existing annotation tool and serves as a basis for further observations.

## CCS CONCEPTS

• **Software and its engineering** → **Interactive games**; • **Computing methodologies** → **Object recognition**; • **Information systems** → Enterprise applications.

## KEYWORDS

gamification, object labeling, training data, machine learning

**ACM Reference Format:**
Sarah Alaghbari, Annett Mitschick, Gregor Blichmann, Martin Voigt, and Raimund Dachselt. 2020. Achiever or Explorer? Gamifying the Creation Process of Training Data for Machine Learning. In *Mensch und Computer 2020 (MuC'20), September 6–9, 2020, Magdeburg, Germany*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3404983.3405519

## 1 INTRODUCTION

Artificial intelligence (AI) is becoming increasingly important. For the development of artificial intelligence, human intelligence is still necessary, especially regarding supervised learning which entails that a machine is trained with labeled data. The training process mimics a human learning process, deriving patterns and creating a model. The creation of necessary labels is usually performed with the aid of humans. Due to the necessary amount of training data the creation process is typically highly repetitive and quickly turns into a rather unexciting, demotivating task for the annotator.

A task that is repetitive and tedious turns out to be the ideal use case for applying *gamification* [19]. *Gamification* itself is defined as the use of game elements in a non-game context [8], aiming

*MuC'20, September 6–9, 2020, Magdeburg, Germany*

ACM ISBN 978-1-4503-7540-5/20/09…$15.00
https://doi.org/10.1145/3404983.3405519

for certain psychological outcomes such as motivation, enjoyment, and flow. Previous research shows that a gamified environment for data annotation has the potential to increase user engagement and gratification [12]. Improved user experience is a goal of gamification, as are increased participation, the attraction of a younger audience, optimization of workflows and increased engagement of users, as well as immediate feedback for the users on their performance [23].

Gamification of company workplaces has just recently gained in importance – not only for the training but also to encourage employees in their daily work routine. A tool with well-designed game elements at the workplace can keep employees motivated to perform their tasks [16]. This paper presents the results of our work aiming at integrating game elements into an existing annotation tool for the creation of training data at the AI product company *AI4BD* [1]. We describe our multi-step development process, thereby laying the foundation for future user studies to investigate the effect of the implemented game elements.

## 2 RELATED WORK

Over the years a lot of different approaches to defining and classifying gamification have been established. During the research, several terms came up which are linked or subordinated to gamification, such as *Serious Games*, *Games with a purpose*, *playfulness*, *gamefulness* and many more. However, the general purpose of gamification is to motivate the target group [8], or more precisely *"using game-based mechanics, aesthetics and game thinking to engage people, motivate action, promote learning, and solve problems"* [13]. Thus, gamification does not mean that a stand-alone game is to be added, hoping for an improvement in employee engagement, but instead to analyze *game mechanics* and visuals, and select game parts which match the use case.

### 2.1 Game Mechanics and Player Types

*2.1.1 Game Mechanics.* We use *Game Mechanics* as the hypernym of *Game Dynamics* and *Game Elements*. By *Game Dynamics*, we denote the strategies and characteristics of games, but also the needs, a player wants to have fulfilled. These needs are, for example, the strive for competition, exploration or social interaction. We will regard *Game Elements* as the actual components found in a game, such as points, leaderboards or avatars. Related literature describes game elements as "the building blocks that can be applied and combined to gamify any non-game context" [15]. This distinction being made, it is still possible to map one to the other. Table 1 shows several *Game Dynamics* as well as *Elements* that trigger them respectively. For example, the *Game Dynamic* progression can be supported by the *Game Element* levels or progress bar, which is intuitively understandable as the feeling of advancing can be triggered with new levels being reached or even unlocked, as well as with a progress bar which is filling up increasingly. The *Game Element* points can be regarded as a progression trigger, under the assumption that the number of points is an indicator for the player's playing skills which means that an increase of points correlates with improved skill. On the other hand, points can also be used to satisfy the need for competition. The dominating motivator in these cases is competence (alternatively called "mastery") [17]. Now, knowing these elements, one might be tempted to simply pick the ones with the greatest appeal and surprise the employees with a generic game layer featuring a leaderboard and random scores. However, this method has its drawbacks and is criticized by [4] who call it the "one size fits all" approach. They suggest a focus on the context which is aimed to be gamified and to consider "specific user needs, goals and values". Therefore, we will follow a user-centered design.

**Table 1: Game Dynamics and suitable Game Elements**

| *Game Dynamic* | *Game Element* |
|---|---|
| Exploration, Surprise [5] | Unlockable [5] |
| Story/Narrative [10] | Story, Badge, Achievement [14] |
| Boundaries [10] | Limited Resources [8] |
| Competition [14] | Leaderboard, Points [14] |
| Resource Acquisition [14] | Achievement [14], Unlockable [5] |
| Status [14] | Leaderboard, Levels [14] |
| Cooperation [14] | Teams [14] |
| Transaction [14] | Gifting [14] |
| Reward [14] | Badge, Achievement [14] |
| Progression [14] | Levels, Progress Bar, Points [14] |

*Source:* The references in the table denote the source used for the mapping between Game Dynamic and Game Element.

*2.1.2 Player Types.* Evidently, the essence of user-centered design lies within the users which is why it is necessary to get familiar with the players. Some authors even recommend a thorough personality analysis of the users, with aid of personality type models such as the Big Five or The Myers-Briggs type indicator [2, 10]. It is assumed that knowing a player's personality traits, gamification can be built according to their personal needs and thus make it easier to trigger their intrinsic motivation. However, the personality type can also provide insight into how prone to certain dangers of gamification a user might be. Several ways of classifying players have been established so far. Notably, many of them are based on Bartle's 1996 theory of four main player types [3]. Bartle's theory was developed based on the question *"What do people want out of a MUD (Multi-User Dungeon)?"*. The author collected players' answers and then categorized them into four main motivations which he turned into player types, meaning classes of users participating in a MUD who share common primary goals. First, there are *Socializers* who aim for inter-player relationships, empathize with people and enjoy observing them. The *Killers* are likewise focused on other players but aim for imposing themselves on others by attacking them and want to win at any cost. The *Achiever* type is also interested in winning but less to defeat others rather than for the sake of points-gathering and rising in levels. Lastly, Bartle defines the group of *Explorers* who enjoy progressive actions, figuring out how things work and the discovery of interesting features. As these player motivations are not mutually exclusive, a real-life player is regarded as a combination of all of these types at different rates, of which some are more and others are less dominant.

[1] AI4BD Deutschland GmbH, https://ai4bd.com/

*2.1.3 Conclusion.* Despite being initially derived from a multi-player game context, Bartle's theory is still highly present in today's player classifications. The names used for the player types can vary greatly. The *Explorer* type, for example, is also referred to as *Free Spirit* or *Creator* [18], *Detective* or *Navigator* [9], depending on the particular focus. What appeals to these explorative players, is a game that is highly adaptable and satisfies their need for *autonomy* with elements such as custom avatars and many unlockable items. Nonetheless, a game with such elements can still attract users of the *Achiever* type who may not willing to spend 30 minutes on choosing an outfit. The possibility to skip such decorative steps should be given in their behalf, as well as additional elements that feed the *Achievers'* competitive needs, such as a leaderboard. A leaderboard might, however, demotivate less competitive users. Therefore, game elements should be selected deliberately and with a lot of attention to the users to prevent unpredicted and undesired behavior.

## 2.2 Gamification for Annotation

Having observed gamification in a general way, we analyzed existing approaches that use game elements in the context of annotation. We present three examples, their setup, features and how they relate to our use case.

*2.2.1 Gamification in video labeling.* A game for video annotation was designed in [21]. They thought out three different game approaches: a label vote game, an entity annotation where users were asked to assign a certain category to a video segment, a click game, where users had to locate a certain object inside the video and click on it, and a bounding box game, which asked users to draw a box around a specific object. The last one was implemented and evaluated with the aid of 20 persons who had not been in touch with the data or the use case ever before. A questionnaire was answered as well, showing that the users liked the game but also agreed that it got more repetitive and boring with time. Used game elements were a progress bar, levels, an optional leaderboard and statistics over experience. The author also mentions the struggle of creating a level system with increasing difficulty for an annotation use case while maintaining the accuracy of the results. In general, the quality of the labels was not satisfying as the resulting bounding boxes were inaccurate. Users also stated they were not willing to spend more time on the tool. Still, the author concludes that a gamified approach could be of advantage concerning annotation cost, given that a very efficient and well-thought-out game is developed. We assume that this is an example of the "one size fits all" gamification approach as apparently, the gamification concept did not regard any adaptive measures towards the user needs. However, this might have been caused by time limits. It has to be mentioned as well, that in this case a full game was developed from scratch, instead of including game elements into an existing tool. Also, unlike our use case, participants were not regularly confronted with an annotation use case and therefore they did not have any experience with this task. We do not see the goal of gamification in convincing every possible user, but in the adaptation and improvement of a tool for a certain group of users.

*2.2.2 Tags You Don't Forget: Gamified Tagging of Personal Images.* Another approach was created by [20] whose scope was the creation of a game, used to annotate personal photos. Two mobile applications were developed (one single, one multiplayer) and evaluated as well as compared to a simple tagging app without any gamification. Concerning game elements, the authors mention that simple playful elements, for example, acoustic feedback for interaction, can already be sufficient in order to motivate a user. The single-player app was a simple tagging app, while the multiplayer app was developed as a Tagger-Guesser-Game. Here, Player B was shown a photo and had to choose between several tags to guess the one that Player A had selected. This approach is similar to the ESP game [22], which uses human aid for image recognition. Assigned labels were rewarded with a point. Correct answers in the multiplayer app were likewise rewarded with one point, while one point was lost for a wrong answer. The labels were evaluated by an expert as being of "good quality". Besides, a questionnaire was answered by the participants to analyze their impressions of the game. They stated that the multiplayer app was much more entertaining, whereas the single-player app helped them memorize the labels better. The insight we take from this example is that less is more when it comes to the selection and amount of game elements.

*2.2.3 Crowdsourcing.* Lastly, we analyzed crowdsourcing tools, which often include game elements to engage users. Google Crowdsource [11] is a desktop platform as well as a mobile app, which makes use of humans to improve Google tools such as Google Photos or Google Translate and can be used by anyone who has a Google account. There are different kinds of tasks that can be performed, e. g., image labeling, approval of image labels, handwriting recognition, and translation, which is close to our use case. In contrast to the two above-mentioned examples, Google Crowdsource is an established user contribution platform which is not a game in itself, but includes game elements, like levels, points, badges, and a leaderboard. It simply works by triggering the basic human needs [17]: *relatedness*, since everything is open and visible, *autonomy* since there is no pressure and users are free to decide when or whether they participate, and above all *purpose* since users get the feeling of being an active part in the improvement of popular software tools.

Another crowdsourcing-based approach incorporating game elements was proposed by Altmeyer et al. [1]. Their goal was to encourage people to keep track of their expenses using OCR (optical character recognition) to analyze grocery receipts. The recognition is trained by crowd input (classifying a given extract of a receipt or categorizing an item). The implemented smartphone application features achievements, points, and a leaderboard to motivate users and to increase the amount of user contribution. However, a dilemma of such crowdsourcing-based approaches is the lack of knowledge about the characteristics of the user group, making it difficult to design for specific user needs and player types.

## 2.3 Dangers of Gamification

As gamification makes use of game elements, it is necessary to keep in mind that with these elements some of their risks might also be adopted. One way to approach this topic has been executed by Callan et al. [6], where ten fictive scenarios of gamification are presented which have been wrongly established in businesses. Recurring problems were a lack of goal-orientation, unsuitable game elements and rewarding, and the danger of revealing too much information to the employees which they might attempt to use for their benefit.

Furthermore, the term addiction is mentioned in this context [2]. Here, however, it is regarded much more as a dependency which users might develop if they get used to the presence of game elements in connection with the task to be performed and hence lose their motivation to perform said task without gamification. As not all possible risks and dangers can be foreseen, another important measure is the constant monitoring of user activities, the detection of abnormalities and suspicious behavior, and a respective adaption of the system [2]. After all, no ideal gamification will be created from the very beginning. Also, no team of employees will stay the same over a longer period, and with people, preferences and needs will change. A promising long-term solution is the creation of an intelligent, adaptive gamification application [2, 4].

## 3 REQUIREMENT ANALYSIS

Deterding et al. [7] describe the procedure of developing a successfully gamified tool as a "full circle" process: *"from formative user research through synthesis, ideation, prototyping, design and usability testing"*. Regarding potential risks of gamification (e.g. wrongly guided motivation, off-task behavior, unwanted competition, addiction and dependency) that might be adopted into a system, it is necessary to define a clear goal that is to be achieved to have a focus while conceptualizing the approach and productive game elements. Concerning the annotation task, we regard three central metrics that can be improved: *quantity* (how many annotations are created), *quality* (how good / correct are the annotations), *enjoyment* (how much fun is the annotation task).

In the following, we first describe the current annotation process with the existing tool support, present our findings from a survey among the employees, and finally sketch two possible gamification concepts for this use case.

### 3.1 Current Annotation Process

The company's existing annotation tool is a multi-user web application prototype which offers registered users a sophisticated annotation environment for collections of images (typically scanned documents). The annotation tasks are of four different types:

- handwriting annotation, where annotators are given an image of a handwritten sequence of letters and numbers which they have to type,
- document classification, where annotators need to classify parts of a document, e. g., to mark tables inside a form using semantic bounding boxes,
- classification, where annotators are asked to identify a given object, e. g., if an image contains a number,
- natural language processing (NLP), where annotators are asked to assign semantic meaning to words, for example, to mark all persons in a given text.

Users can see all the collections which are assigned to them, including their annotation state, i. e., how many of the items within the collection were approved, annotated and refused (i. e., rejected because it was too ambiguous). Selecting a collection, users can see a grid view of the contained resources, color-coded depending on their state (grey: open, blue: approved, red: refused). Additionally, users can also see the number of annotations that have been created for each resource. Selecting a document, users enter the annotation view itself, where they can create an annotation in case the document's state is open, or see its state and annotations that were created for this resource. In case users are insecure and wish to access annotation guidelines, they either need to navigate to an external annotation tutorial or ask their coworkers.

In order to ensure quality, an annotation is being reviewed after the creation. Selected users who have the role "reviewer" assigned to them can access additional features in the annotation view allowing them to approve or refuse an annotation. The review of handwriting annotations is currently semi-automated by automatically marking an annotation as "approved" if at least two distinct annotators create an annotation with the same value.

### 3.2 User Survey

We conducted a user survey among company employees working as *annotators*, to get an idea of their characteristics, whether a gamified approach would appeal to them at all, which game elements would suit them most, and which should be avoided regarding the aforementioned potential risks.

We adapted the "student model" from the work of Andrade et al. [2], which defines five attributes of the player: *Knowledge*, *Psychology*, *General Behavior*, *Gamer Profile*, and *Interaction*. As General Behavior is focused on personal habits unrelated to the domain, we decided to omit this due to privacy issues. The Interaction attribute addresses information about the user activities which is better obtained via monitoring and logging (e. g., number of logins, success rate). We also decided to leave this out as it was not our goal to assess individual user activity.

Consequently, we created a questionnaire covering the three aspects *Knowledge* (labeling experience), *Psychology* (personal opinions) and *Gamer Profile* (game experience). Twenty company employees participated in the survey (11 of them aged between 24 and 30, two younger than 24, four aged between 31 an 40, one older than 40, two preferred not to tell their age). The only mandatory question was if they had ever performed an annotation task. If they had, they could answer more follow-up questions referring to annotations. All other questions were voluntary.

*3.2.1 Knowledge.* When asked about their experience, 18 out of the 20 participants stated that they have already performed annotation tasks for the company, half of them indicated that they have been labeling data for more than three months. In a multiple-choice question, we asked the 18 participants who had experience with annotation which kinds of labeling tasks they had already performed. Document placement (15) and handwriting recognition (14) were the ones that had been performed by most of the annotators, followed by NLP tasks (8) and classification (6).

*3.2.2 Psychology.* Concerning the psychological aspect, we asked the annotators to take a position on six moderately provocative statements, choosing from a Likert scale of five different options of agreement (*I agree... not at all (-2) / not quite (-1) / neutral (0) / a bit (1) / a lot (2)*). From the number of positive answers (Likert scale values 1 and 2), we derived a percentage for the agreement per statement.

- *"I find labeling tasks tiresome"* (65% agreed, M=0.7, SD=1.117)
- *"I would like to be able to see how well I am doing in labeling, compared to my coworkers"* (55% agreed, M=0.2, SD=1.348)
- *"If labeling included game elements, the label results would be better"* (50% agreed, M=0.4, SD=0.993)
- *"If labeling included game elements it would be much more fun"* (65% agreed, M=0.9, SD=0.999)
- *"I would not like it if others were able to see my labeling progress on a leaderboard"* (45% agreed, M=0.35, SD=1.27)
- *"Using game elements at work makes a company seem less serious"* (30% agreed, 55% disagreed, M=-0.65, SD=1.306)

From these results we derived the following conclusions: While labeling is generally considered rather tiring, the score for the statements encouraging game elements were overall positive and game elements are agreed upon as a promising tool for making labeling more fun without harming the image of the company. However, most of the annotators dislike the idea of being able to see their labeling progress on a leaderboard, but do not generally object to compare themselves with colleagues.

*3.2.3 Gamer Profile.* As for the gaming habits, we posed questions regarding the time spent on games, which kinds of games were preferred as well as which game elements were the most motivating ones. The majority likes digital games, with 60% of them playing them at least every week, whereas 20% played them at least once a month and 10% only rarely or not at all, respectively. We added the question on real-life games, in case the participants were not keen on digital games, but still liked playing physically. In our group, however, digital games were more popular.

To figure out which of Bartle's four main player types [3] was most present in the study group, we asked the participants to indicate how much they enjoy distinctive game types (like simulation games, action games, puzzle-based games, etc.), using a five-level Likert scale (*not at all*(-2) ... *a lot* (2)). Furthermore, they were also asked to rate specific game elements (like leaderboards, playing against others, rewards, team play, etc.) on a five-level Likert scale (*very demotivating (-2) ... very motivating (2)*). Finally, we also asked about the dominating motivation to play games at all. We used the correlation between preferred game types, preferred game elements and gaming motivations to create a score for each set of player type characteristics per player. The resulting scores are shown in Figure 2. We identified nine participants with predominant characteristics of an *Achiever* (points-gathering, rising in levels), six more inclined to be an *Explorer* (progressive actions, find interesting features), and two showing equal characteristics of both (P9, P15). Thus, a group tendency towards Achiever and Explorer characteristics was notable.

*3.2.4 Further Feedback.* We also asked the annotators what they disliked about the current tool and how they would like it to be improved. From this information, we hoped to be getting some impressions of possible stimuli for a gamification concept. Except statements like *"Labeling is boring."*, the feedback we got for this question was concerning more technical issues, like the repeated demand for a more fluid user interaction inside the annotation tool by supporting key shortcuts to reduce the need for mouse interaction.

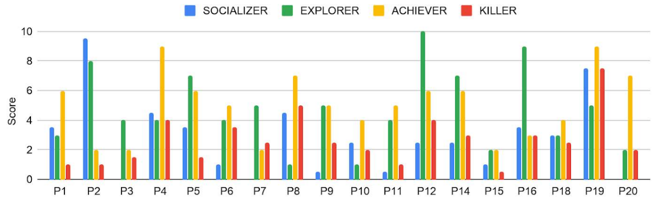


**Figure 2: Player type characteristics of each participant. The score is based on the answers regarding preferred game types, preferred game elements and gaming motivation. Note: P13 and P17 are excluded here as both considered *all* elements to be "demotivating" (negative score).**

*3.2.5 Lessons learned.* From our survey, we can conclude that the majority of annotators are playing games and are open to the use of game elements inside work tasks. We learned that a way of comparison of performance is desired, but should provide anonymity, which is also a precaution in terms of the danger *Unwanted Competition*. A complex narrative, levels with increasing difficulty, as well as playing with others in a team, but also playing against others and exploration were voted as the most motivating game elements. The dominating player type characteristics in the group of annotators hence turned out to be the ones of the *Achiever* and the *Explorer* type. Based on these findings, we created a gamified concept for an annotation tool, described in the following.

## 3.3 Basic Gamification Concepts

We combined the results of the survey into two different concepts, each containing a selection of the preferred game elements:

1. *I can make a change:* Story / narrative, levels, progress bar, badge / reward (*Explorer* type)
2. *TeamChallenge - Us versus them:* Competition between teams, leaderboard with team names, points and achievements (*Achiever* type)

We presented both basic concepts to the annotators themselves, giving them a chance to give feedback and express their opinions concerning the idea of having said game elements inside their tools. While the competitive approach seemed appealing, the first one of building a story around the annotation tasks was preferred. Furthermore, the idea of incorporating a narrative led to much valuable creative input on the part of the annotators. Some ideas which were named were to divide the big narrative into various chapters and steps (hence levels) that need to be passed, but also that the story - if it is told correctly - could help the annotators understand what their work was used for and why they were doing it. So subconsciously, they expressed the desire for *Purpose* (people find a task much more intriguing when there is a reason to do it or a greater meaning behind it, which can also be linked to altruism according to [17]) .

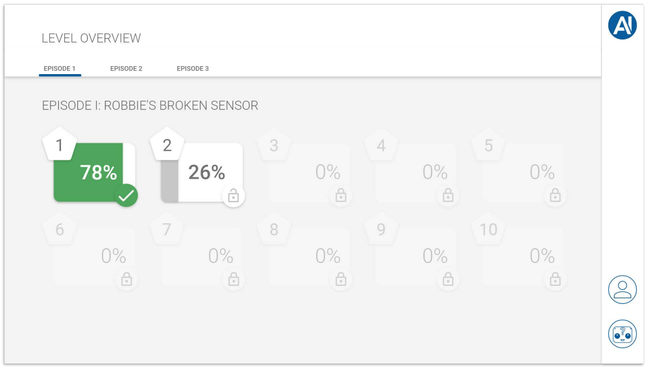


**Figure 3: The central level overview, with a header showing the episode tabs and a sidebar with the menu points "my statistics" and "help".**

Further, it was mentioned that despite not implementing a team feature per se, the tool could still support the group feeling by showing the general group progress for all annotations. Related literature refers to this aspect as *"Perceived visibility"*, being *"related to the notion of being noticed by peers and in a position of social presence"* [2]. On the other hand, concerns were voiced as to how motivating such a progress bar would be in case there would be a long period without any change or if the model was adapted, causing the progress to decrease. However, the story element leaves a lot of room for adaptations in this case, e. g., negative progress could be explained using a negative twist inside the narrative itself. This requires a constant monitoring of user performance and a thorough player model to detect anomalies and have the system react in such a way that user behavior is directed correctly.

## 4 GAME DESIGN AND IMPLEMENTATION

From the discussion with the users, we derived a final game concept that includes a complex narrative, levels of increasing difficulty, and a progress indicator.

Upon first use, the annotators are taken on a tour through the tool where all the central elements are described, and the tasks and the narrative are introduced. Knowing the basic story and the main goal, users get to the level overview (Figure 3) which from then on is always going to be the initial screen. Users can switch between different episodes that each can be used to tell different storylines, but also for annotation tasks of different types. New episodes can be added by administrators (or authors) at any time, so they do not depend on the user's progress. Clicking on one level, the users enter the level screen where the different resources which are assigned to this level are listed (see Figure 1, left). From here, they can choose a resource to annotate (see Figure 1, right). The following subsections are each dedicated to one game element, describing its design, placement, and function.

### 4.1 Story

In order to support intrinsic motivation, a story element for gamification should be linked to the context and the tasks performed by the company [2]. Hence, we got inspired by Google Crowdsource [11] where the speech assistant training task is initialized with a robot which introduces the topic to the user. We created the AI user assistant "Robbie" to fulfill three jobs: it is a tutor which gives first-time users a tour through the tool, it is the *"Help"* element of the tool and it is the center of our narrative, telling the story and giving feedback on the progress. Each episode has one main storyline where Robbie faces a struggle that needs to be solved with the help of annotations. Examples for these stories are to help Robbie achieve certain capabilities, such as learning the human language, which includes learning to "read" (handwriting annotation), to "understand" document structures (bounding box tasks) or even linguistic structures (entity annotation). In the following levels, these plots can always be reused by asking the user to train Robbie further in terms of one of the mentioned capabilities. With these plots, we aim to support the user need *purpose* by creating abstract stories that are related to the real-life use case.

### 4.2 Levels

Inside a level, users can see all of the resources, including the ones annotated by other users. For this reason, there is a filter bar which annotators can use to filter the resources by their state and by "only my resources" (switch-toggle button). Additionally, each resource which was annotated or already started by this user has a user icon in the top right corner. In the level head, users can see the level's quest in the center and the number of their current score inside this level in the right corner. This score shows the number of approved or the number of made annotations (depending on the quest) out of the number of annotations needed to pass the level. The quest itself is a narrative element. It asks the user to reach a certain annotation goal which leads to fulfilling a greater purpose (of helping Robbie). The way the quest is phrased and designed is essential for the fulfillment of the gamification goal. For now, we distinguish two basic quest types: *"Annotate a certain number of resources!"* (quantity) or *"Get a certain number of approvals for your annotations!"* (quality). The main goal of our gamification approach is to improve the quality of the annotations. Consequently, for the most part, our quests will require users to create *approved* annotations or combine both quest types (*"Annotate x resources and get y approvals!"*).

### 4.3 Progress

In order to keep track of the quest realization, the user's progress needs to be visualized. Progress is shown in the level overview, where users can see what percentage of the total resources inside the level has already been annotated and approved and if they passed the level or not. Furthermore, the possibility of unlocking new levels if the user passed a level is another user-specific progress element. Inside each level, there are multiple progress elements: the annotation counter in the level head, the resource colors, the level theme image (which changes from greyscale to color upon passing a level), and the level progress bar. In the following sections, we explain in detail the layout of this progress bar and how we adopted colors into the progress concept.

*4.3.1 Colors.* We use four main colors for encoding progress (Figure 4). According to their state, the background color for a level box in the level overview and a resource in the level view is determined.

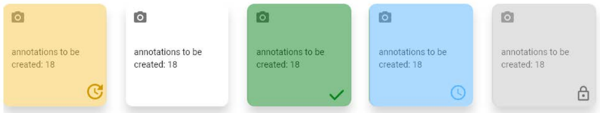
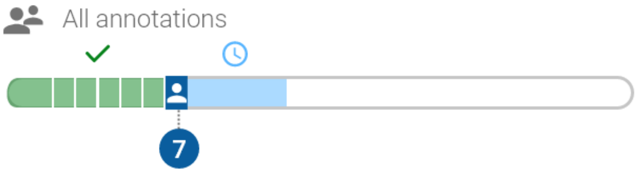

annotations to be created: 18
annotations to be created: 18
annotations to be created: 18
annotations to be created: 18
annotations to be created: 18

**Figure 4: The different colors of resources, depending on their states yellow: rework, white: open, green: approved, blue: in review, grey: locked**



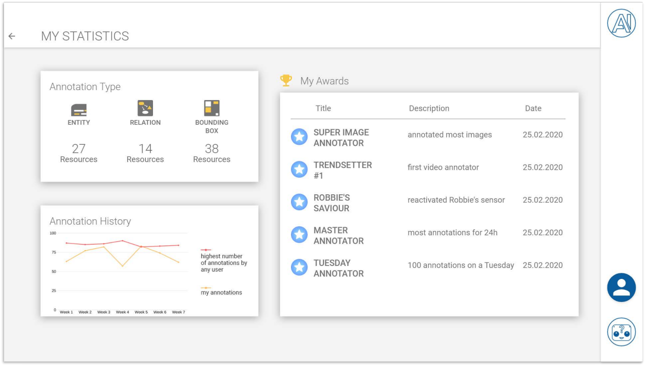
You reached the seventh highest number of approved annnotations in this level.

**Figure 5: The progress bar shown inside the level view**

Green represents the state *passed* (level) or *approved* (resource). White represents the state *open* and is used for the background of a level box that contains open resources and of an open resource which still needs to be annotated. Grey denotes the state *locked* and is used for levels that are not accessible yet as well as for resources that can not be accessed because they are currently being annotated by another user. A resource has two additional potential states: being in review after a user finished annotating it (which is shown with blue color), and being in rework if a resource has been reviewed and not approved due to incorrect or insufficient annotations. The same color palette is used for the progress bar inside a level.

*4.3.2 Progress Bar.* The progress bar contains information on how many annotations in this level have been approved (green) and how many annotations are in the review process (blue), in proportion to the whole number of resources. The remaining white space of the bar implicitly encodes all open resources inside this level. The green space is additionally divided into several parts, each one representing a user and their approved annotations. These partitions are sorted from left ("most approved annotations") to right ("least approved annotations") and they are anonymous, so no user can see which one belongs to whom. They can, however, see where their part is located which is highlighted in blue and with a user icon (Figure 5). So, apart from serving as progress information, the bar also serves as an anonymous leaderboard.

*4.3.3 Success Notifications.* When a level is passed and the next one gets unlocked, a user will get a success notification. Where and when exactly it appears, depends on the passing rule. If the quest goal is to create a certain number of annotations only, the success notification will appear inside the annotation tool, showing a happy Robbie celebrating and giving the user the options to go to the next level or proceed to create annotations for the current level. If the passing rule demands a certain number of approvals, the moment when a level is passed does not correlate with the annotation flow or even with the time when the user is online. In this case, the success message will appear in the level overview, with an animation showing the next level being unlocked. Alternatively, a pop-up notification inside the tool can be shown if the user is online when the required number of annotation approvals is reached.



| Title | Description | Date |
|---|---|---|
| SUPER IMAGE ANNOTATOR | annotated most images | 25.02.2020 |
| TRENDSETTER #1 | first video annotator | 25.02.2020 |
| ROBBIE'S SAVIOUR | reactivated Robbie's sensor | 25.02.2020 |
| MASTER ANNOTATOR | most annotations for 24h | 25.02.2020 |
| TUESDAY ANNOTATOR | 100 annotations on a Tuesday | 25.02.2020 |

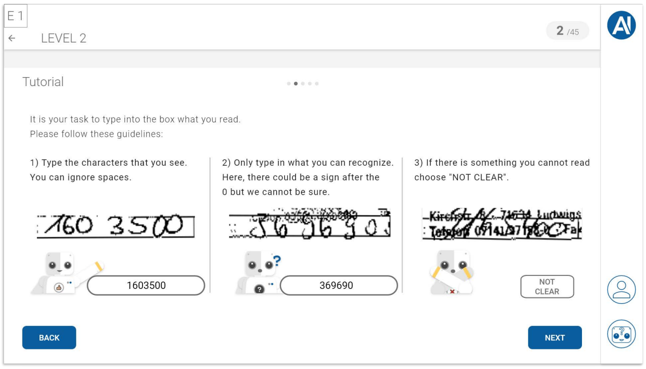
**Figure 6: A user’s personal statistics page showing achievements, number of annotations created per annotation type, as well as a history chart showing total number of annotations over time.**

## 4.4 Statistics

In the sidebar, users can access a general help screen by clicking on the Robbie icon, but also access their personal user statistics. This is a feature we proposed due to the fact that users did like the idea of seeing their progress, also in comparison to others as derived from the results of the user study. In the statistics screen, shown in Figure 6, they can see awards they got, the number of annotations they created per annotation type, as well as a chart showing the history of their total number of annotations. The ideas for these charts are just an initial suggestion and can be adapted to the users’ needs.

## 4.5 Tutorial

Users see a tutorial after they click on an open resource if this requires annotation of a type which the user is not experienced in or in case the resource has any kind of exceptional rules which the annotator must know. Currently, the company has tutorials that are not embedded inside the annotation tool. By providing this feature, we aim to support the improvement of the annotation quality as it requires users to read the rules and guidelines before creating the annotation. The tutorial consists of multiple steps: first, the rules are presented, with Robbie as a decorative element, highlighting positive rules (what is recommended, examples for a good dataset) and negative rules (which data should be skipped), as shown in Figure 7. They are followed by a brief training part where users get confronted with minimal annotation tasks that test their understanding of the previously presented rules.



It is your task to type into the box what you read.
Please follow these guidelines:

1) Type the characters that you see. You can ignore spaces.

2) Only type in what you can recognize. Here, there could be a sign after the 0 but we cannot be sure.

3) If there is something you cannot read choose "NOT CLEAR".

**Figure 7: A tutorial page, introducing an annotation task with positive and negative rules and recommendations.**

## 4.6 Working Prototype

Starting with low-fidelity prototypes, we finally implemented a fully operable Web prototype within the company’s annotation environment, based on Angular 9 and NgRx, and using a central backend service, which obtains data from a MongoDB database. In addition to the former annotation tool, the prototype introduces software components to represent episodes, annotation levels and awards as game elements. The prototype is fully operational regarding real annotation tasks, user authentication and login, as well as loading and saving annotations per user account. Thus, preexisting and gamified annotation tool are ready for use, e. g. to be compared in long-term A/B tests. Short-term tests have already been conducted to verify functionality, but these are not yet meaningful regarding our four central metrics, *quantity*, *quality* and *enjoyment*. Nevertheless, initial feedback from the company and individual employees on the results was throughout very positive and encouraging.

## 5 CONCLUSION AND FUTURE WORK

This work describes our approach and design process for the gamification of an annotation tool for creating machine learning training data. Unlike a "one size fits all" approach to gamifying the tool, where game elements are applied without regard to the context of use [4], we choose a user-adapted approach which first analyzes existing literature for gamification and then performs a user research study with twenty employees of the company. The results show that the employees enjoy gaming in their free time, which supports the utilization of a gamified tool, and which game preferences they have. From the findings, we derive an individual gamification concept with regard to the annotation use case and the employees’ player characteristics. Our implemented prototype is a gamification approach for the company’s annotation tools. It serves as a proof of concept that game elements can be easily implemented inside an existing environment.

One aspect we did not cover in this work is how to assess the difficulty of an annotation task. One approach is to map the complexity, hence the effort, of the task to the difficulty. On the other hand, even a less complex task can be of greater effort than a complex task if the data contains a lot of ambiguousness. Future work can analyze this problem thoroughly. It might also be helpful to follow a more thorough approach of user research that considers psychological aspects, possibly even personality types, and aims for a deeper user analysis. Generally, it can be interesting for other projects to regard other taxonomies of player types and to perform detailed psychological user research. Besides, we did not evaluate our approach over a long period (over several months), which is especially necessary when a narrative is included which is an element that evolves.

During our research, we found a variety of different taxonomies for gamification. We also noticed that not all terms are used consistently by different sources, for example, the terms Game Mechanics and Game Elements. Besides, many approaches to distinguish player types exist, which is why we chose to stick with the basic player type taxonomy by [3]. Researchers with a similar purpose should keep in mind that gamification is perceived with skepticism and concerns by some users as the underlying idea is the manipulation of user behavior. For the prevention of a negative impression, it is recommended to include the users in the design process by asking for their opinions and their general game affinity. We strongly discourage any destructive intentions when using gamification, for example, aiming for surveillance of the staff or a highly competitive environment in the company. Related work also frequently mentions the importance of transparency and disclosure concerning the game tool. One possible way to encourage trust is by giving the users access to information on the reasons for the use of game elements and not leaving them with a wrong feeling of being observed or put under pressure by a gamified tool.

## ACKNOWLEDGMENTS

This work was partially funded by the German Research Foundation (DFG, Deutsche Forschungsgemeinschaft) as part of Germany’s Excellence Strategy – EXC 2050/1 – Project ID 390696704 – Cluster of Excellence "Centre for Tactile Internet with Human-in-the-Loop" (CeTI) of Technische Universität Dresden, and by DFG grant 389792660 as part of TRR 248 (see https://perspicuous-computing.science). Additionally, the work at AI4BD was funded by the German Federal Ministry of Education and Research and the research project PANDIA (grant number 16SV8396).

## REFERENCES

[1] Maximilian Altmeyer, Pascal Lessel, and Antonio Krüger. 2016. Expense Control: A Gamified, Semi-Automated, Crowd-Based Approach For Receipt Capturing. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (Sonoma, California, USA) *(IUI ’16)*. Association for Computing Machinery, New York, NY, USA, 31–42. https://doi.org/10.1145/2856767.2856790

[2] Fernando Andrade, Riichiro Mizoguchi, and Seiji Isotani. 2016. The Bright and Dark Sides of Gamification. In *Intelligent Tutoring Systems*, Alessandro Micarelli, John Stamper, and Kitty Panourgia (Eds.). Lecture Notes in Computer Science, Vol. 9684. Springer International Publishing, 1–11. https://doi.org/10.1007/978-3-319-39583-8_17 06.

[3] Richard Bartle. 1996. Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD research* 1, 1 (1996), 19.

[4] Martin Böckle, Isabel Micheel, Markus Bick, and Jasminko Novak. 2018. A Design Framework for Adaptive Gamification Applications. In *Proceedings of the 51st Hawaii International Conference on System Sciences (Proceedings of the Annual Hawaii International Conference on System Sciences)*, Tung Bui (Ed.). Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2018.151

[5] Bunchball. 2015. *What are Game Mechanics?* https://www.bunchball.com/gamification/game-mechanics

[6] Rachel C Callan, Kristina N Bauer, and Richard N Landers. 2015. How to avoid the dark side of gamification: Ten business scenarios and their unintended consequences. In *Gamification in education and business*. Springer, 553–568.

[7] Sebastian Deterding, Staffan Björk, Lennart Nacke, Dan Dixon, and Elizabeth Lawley. 2013. Designing gamification: creating gameful and playful experiences. In *Extended Abstracts on Human Factors in Computing Systems*. 3263–3266. https://doi.org/10.1145/2468356.2479662 04.

[8] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining "Gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (New York, NY, USA) *(MindTrek '11)*. ACM, 9–15. https://doi.org/10.1145/2181037.2181040

[9] Dominique Mangiatordi. 2018. *Gamification at work: the 8 PLAYER TYPES*. https://www.linkedin.com/pulse/gamification-work-8-player-types-dominique-mangiatordi-/

[10] Lauren Ferro, Steffen Walz, and Stefan Greuter. 2013. Towards personalised, gamified systems: An investigation into game design, personality and player typologies. In *Proceedings of The 9th Australasian Conference on Interactive Entertainment Matters of Life and Death*. https://doi.org/10.1145/2513002.2513024 09.

[11] Google. 2020. *Google Crowdsource*. https://crowdsource.google.com/

[12] Simone Hantke, Tobias Appel, and Björn Schuller. 2018. The Inclusion of Gamification Solutions to Enhance User Enjoyment on Crowdsourcing Platforms. In *2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*. 1–6. https://doi.org/10.1109/ACIIAsia.2018.8470330 05.

[13] Karl Kapp. 2012. *The gamification of learning and instruction: Game-based methods and strategies for training and education. San Francisco, CA: Pfeiffer.*

[14] Selay Arkün Kocadere and Şeyma Çağlar. 2018. Gamification from player type perspective: A case study. *Journal of Educational Technology & Society* 21, 3 (2018), 12–22.

[15] Janaki Kumar and Mario Herger. 2013. Gamification at Work: Designing Engaging Business Software. In *Design, User Experience, and Usability. Health, Learning, Playing, Cultural, and Cross-Cultural User Experience* (Berlin, Heidelberg), Aaron Marcus (Ed.). Springer Berlin Heidelberg, 528–537.

[16] Daria Lopukhina. 2018. *How Gamification in the Workplace Impacts Employee Productivity*. https://anadea.info/blog/how-gamification-in-the-workplace-impacts-employee-productivity

[17] Andrzej Marczewski. 2013. *The Intrinsic Motivation RAMP*. https://www.gamified.uk/gamification-framework/the-intrinsic-motivation-ramp/

[18] Andrzej Marczewski. 2015. *Even ninja monkeys like to play: Gamification, game thinking & motivational design*. Gamified UK. http://gamified.uk/user-types/

[19] Jane McGonigal. 2011. *Reality is broken*. Penguin Press. http://www.loc.gov/catdir/enhancements/fy1107/2010029619-d.html

[20] Nina Runge, Dirk Wenig, Danny Zitzmann, and Rainer Malaka. 2015. Tags You Don't Forget: Gamified Tagging of Personal Images. In *Entertainment Computing - ICEC 2015* (Cham), Konstantinos Chorianopoulos, Monica Divitini, Jannicke Baalsrud Hauge, Letizia Jaccheri, and Rainer Malaka (Eds.). Springer International Publishing, 301–314.

[21] Samuel Suikkanen. 2019. Gamification in video labeling; Videoiden merkkaamisen pelillistäminen. http://urn.fi/URN:NBN:fi:aalto-201906233997

[22] Luis von Ahn and Laura Dabbish. 2004. Labeling images with a computer game. In *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04* (New York, New York, USA), Elizabeth Dykstra-Erickson and Manfred Tscheligi (Eds.). ACM Press, 319–326. https://doi.org/10.1145/985692.985733

[23] Lincoln Wood and Torsten Reiners. 2015. Gamification. In *Encyclopedia of Information Science and Technology*. 3039–3047. https://doi.org/10.4018/978-1-4666-5888-2.ch297 01.