ETH zürich

Media Technology Center

Media Technology Center

Annual Report 2022

MTC D INFK

“It is key to utilise the immense technological progress for the benefit of journalism.“

Dr. Pietro Supino

We live in exciting times: The media industry is in full transformation. I am convinced that we will see more change in the next five to ten years than that experienced so far. We currently face many complicated challenges, but there are even more opportunities to still be explored. In order to capture these, we need to adapt constantly to the evolving circumstances and be innovative. We must continue with our ambition to offer our users the best possible media in the future. Our purpose is reaching, informing and inspiring people and thereby contributing to a free society.

At a time of rapid change in how people consume news and information, the media must adapt to its new realities. Numerous steps have already been taken to move forward, but change shall not stop, it must go on.

It is key to utilise the immense technological progress for the benefit of journalism. It has led to the disruption of old business models. But it also offers many possibilities to create new values and to become more efficient. The use of data analysis, for example, has brought enormous development: it enables the collection, analysis, understanding and explanation of complex contexts in ways that were previously not imaginable. Also, journalism has become much faster. Regular updates of digital channels are required. With this speed and the available enormous volume of information, fact checking and verification are becoming ever more important. Trustworthiness and credibility are the capital and the essential quality of solid media companies. Here too, technology can help to fulfil the task and to make the difference.

The development has been particularly dynamic last year in the field of natural language processing, which is a strength of the MTC under the intellectual leadership of Prof. Ryan Cotterell. The underlying use of artificial intelligence offers in general new starting points for the reinvention of journalism. Again, it brings new opportunities, but also threats and potential pitfalls.

Today and in the future, media is the combination of journalism and computer science. Only this interdisciplinary collaboration allows us to overcome the unfolding challenges and to capitalise on the new possibilities. Beyond the necessity, it is enriching and stimulating to open up the mind to the combination of these complementary skill sets in order to explore new paths and make the world a better place to live in.

In this context the MTC has a very important role to play. By linking one of the world's foremost academic institutions and the media industry, it builds a pool of experiences, knowledge and creativity for the development of ideas and solutions. It creates the required critical mass of a small country in order for it to be in the international forefront of development, and it helps to attract new talent to the media industry. I am proud to be part of this venture and would like to thank and to congratulate all colleagues for their engagement!

Dr. Pietro Supino
Publisher and Chairman, TX Group

People

Core Team

Alberto Pennino

Armando Schmid

Ayça Takmaz

Clara Fernandez

Daniel Vera Nieto

Fábio Porfírio

Fabio Zünd

Julien Heitmann

Laura Mascarell

Majed El Helou

Petar Stamenkovic

Philippe Schlattner

Ribin Chalumattu

Saikishore Kalloori

Severin Klingler

Thomas Steinmann

Industry Experts

Cristina Kadar
NZZ

Olivier Bühler
SRF

Isabelle Schrills
Ringier AG

Stefan Wabel
Verband Schweizer Medien

Dominic Herzog
TX Group AG

Academia Experts

Prof. Andreas Krause

Prof. Ce Zhang

Dr. Donald Tillman

Prof. Markus Gross

Prof. Martin Zimper

Prof. Olga Sornike-Hornung

Prof. Otmar Hilliges

Prof. Ryan Cotterell

Prof. Thomas Hofmann

› MTC's steering committee, composed by industry and academia experts, is the final authority for all decisions regarding the center.

German News Summarization

Summarization plays an important role in the journalistic domain. In 2022, the MTC implemented single- and multi-document approaches to German summarization and a demo application that supports the task of writing the lead paragraph of an article.

Automatic text summarization is the task of automatically generating a shorter version of a document that preserves its most important information. Some tasks in the journalistic domain, such as writing the lead paragraph of an article, would benefit from using a summarization system that generates a summary of the most important facts in the article. State-of-the-art automatic summarization systems are able to generate fluent and coherent text. However, they present some limitations, especially for languages other than English, such as German, which lack dedicated datasets. In particular, summarization models are prone to hallucinate facts that are not supported by the source document.

Single-document summarization with guidance

The lead text of a news article aims at capturing the reader's attention and provides them with the most important information. It usually sets the tone for the rest of the article and gives readers a sense of what to expect. Writing the lead text is in fact an important task that is mostly done manually, since there are no dedicated models for this task. We therefore built single-document summarization models to assist journalists in this process. Specifically, our models implement the use of guidance (e.g. sentences or keywords that highlight the important information in the text) that the user can provide to steer the summary towards it. We have been working with TX Group, NZZ, and SRF to train our models for German text. In addition, the MTC has been implementing a project demo application that will help journalists to write the lead paragraph (see Figure 1). In other domains, such as machine translation, post-editing has been adopted as a more efficient solution than performing the task from scratch. Our web tool provides the journalists with an automatic summary of the article that they can post-edit if necessary. The tool also allows them to highlight the important information from the source article in the user interface, which our model will use to guide the generation of the summary.

Low-resource summarization approaches

When a special event occurs, it is common for different newspapers to report on it through various articles. Automatic summarization of multiple related articles would help the reader to get a comprehensive understanding of a particular event (see Figure 2). However, generating a summary from a set of articles is a challenging task. While neural encoder-decoder models have shown promising results for single-document summarization,

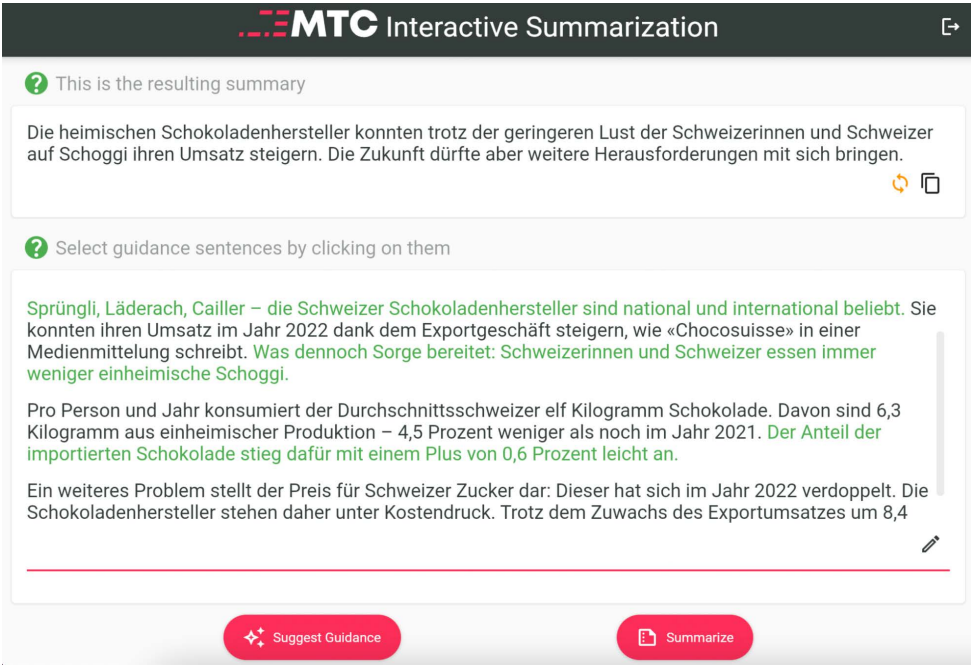

Figure 1: screenshot of the project demo interface that supports the task of writing lead texts. The sentences in green are manually selected to highlight the important information in the source article, which are then used as guidance to generate the summary. After the summary has been generated, the user can post-edit the resulting text if necessary.

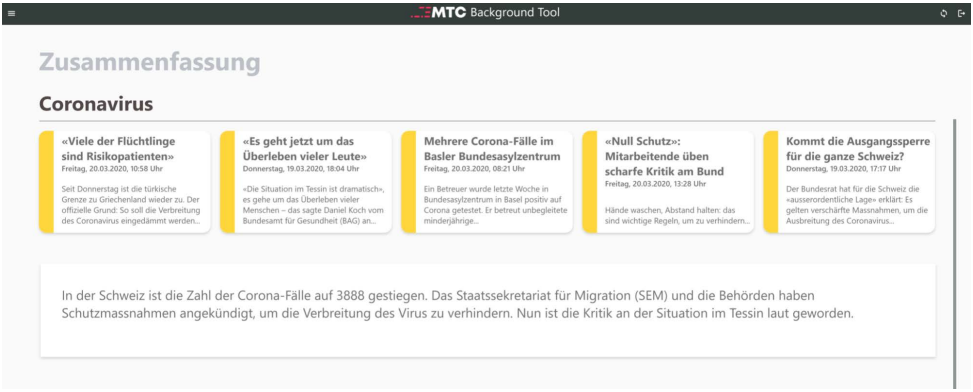

Figure 2: multi-document summarization of German news articles.

they were trained on large datasets automatically extracted from the internet. In contrast, multi-document summarization datasets are very scarce, mostly available in English, and very costly to obtain. We therefore focus on low-resource approaches to multi-document summarization. In particular, we implemented several multi-document summarization approaches that leverage single-document summarization models for German text. We are planning to evaluate their quality in a study together with journalists from SRF.

Faithfulness and Factuality in Summarization

A limitation of the state-of-the-art abstractive summarization models is that they cannot ensure that the summaries are factually consistent with the input document. This poses a major issue in the journalistic domain, since it is essential to provide the reader with accurate information. Additionally, there is a lack of research on faithfulness and factuality in summarization of German text. In contrast to English, there is no metric to evaluate hallucinations in German. Existing approaches mostly use manual annotations on English summaries and/or adapt existing models from other tasks (e.g. textual entailment or question answering) trained on English datasets, making them unusable in other languages.

Recently, the Rycolab research group at ETH proposed a new approach to mitigate hallucinations that leverages pointwise mutual information during decoding. The proposed method can be applied to other languages and we adapted and integrated it in our multi-document summarization approach. To assess its performance, we are conducting a human evaluation following the guidelines of the related work. More specifically, we split the summaries into text spans and ask the annotators to indicate whether the fact in each text span is supported by at least one of the source articles (see an example of the layout in Figure 3). The annotation of faithfulness in summarization is considered a challenging task and the evaluation of spans instead of full summaries helps to improve inter-annotator agreement. ■

Figure 3: Annotation layout of faithfulness in summarization. The text spans are extracted from the multi-document summarization systems that we are evaluating.

- Our web-based summarization assistant supports journalists in writing the lead text of news articles.
- Our models are trained and evaluated on German newswire datasets from NZZ, TX Group, and SRF.
- Our single-document summarization models trained on German data support the use of sentence or keyword guidance to emphasize the importance of specific contexts.
- Our multi-document summarization approaches do not require the use of multi-document summarization datasets for training and outperform the quality of the state-of-the-art models for German text.

Exploring Emotions and Stance in German Text

Machine learning models have greatly improved the estimation of emotion and stance in English texts. However, these models have not been fully explored for German texts. To address this, we created a high-quality labeled dataset for German texts, built tools to visualize emotions and stances in a given text, and we explored the use of emotions in a recommender system.

An important field in natural language processing is the automatic extraction and analysis of information from text, which addresses a variety of research issues including sentiment analysis, truthfulness and rumor identification, emotion recognition, or false news detection. The task of emotion and stance detection can help to better curate produced content from journalists and users, and can allow for a better understanding of reactions, behaviors, opinions, and feelings of readers towards produced content. Moreover, the information about the stance or emotions conveyed in a text can provide additional insights into user engagement metrics, help with estimating propensities of users to register, or subscribe, or potentially even open new opportunities in advertising. While technologies for estimating the emotion and stance in text for English have seen large advances in the past years, these technologies lack behind for texts in German. In this project, the goal is to close the performance gap between methods for English and methods for German.

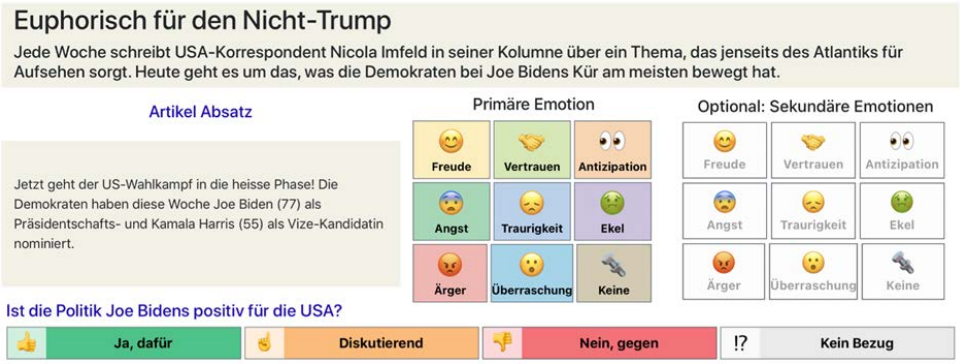

Figure 3: The annotation tool used for creating the dataset.

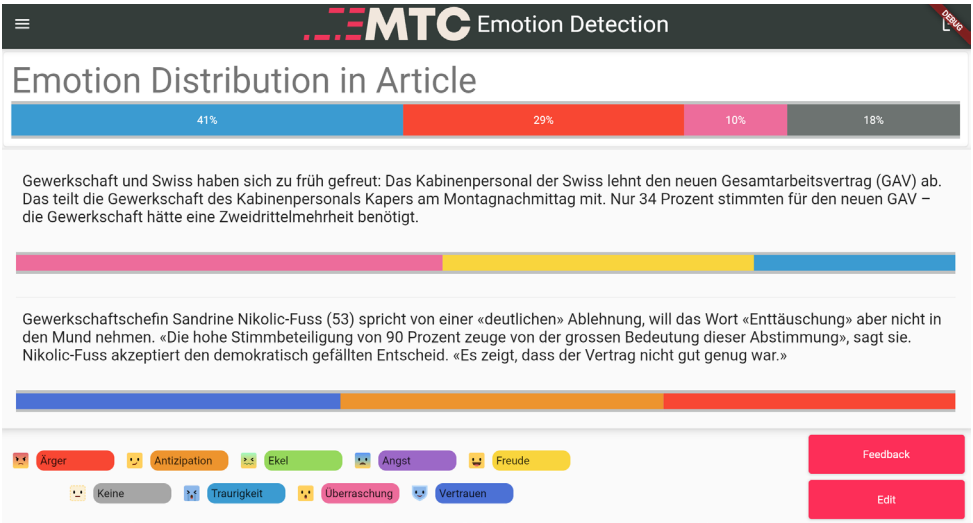

Figure 4: The emotion viewer tool.

Stance and Emotions in German Text

Stance detection is the task of predicting the position of a text (e.g., in favor or against) towards a specific question, which has shown to aid at identifying the veracity of rumors (rumor classification) and fake news. We consider the German counterpart of the English 'in favor', 'against', 'discussion', and 'unrelated' as the stance labels, where 'discussion' indicates that the article discusses the topic of the question without taking a position. For emotion detection, the task is to predict for a given text from a news article, the emotions that the text conveys. We consider a set of eight emotions labels, which are the German counterpart of the basic emotions defined in Plutchik (1980)'s model: Freude ('joy'), Vertrauen ('trust'), Angst ('fear'), Antizipation ('anticipation'), Traurigkeit ('sadness'), Ekel ('disgust'), Ärger ('anger'), and Überraschung ('surprise'). The 'no emotion' label is defined as the German Keine.

Annotation Tool and Dataset

We have developed an annotation tool that can enable the annotators to provide a stance label for a given question-articles and an emotion label for a given article text on paragraph level and full-article level. Our annotation tool (see Figure 3) has simple UI and functionality to elicit the emotions and stance labels. To perform this annotation task, we recruited a team of 23 native German speakers and used our built annotation tool for this purpose. In particular, the annotators, 15 women and 8 men, are university students or academics between 21 and 49 and a median of 25 years old.

Using the data of ~670k news articles provided by the Swiss media companies Blick (~536k articles), Neue Zürcher Zeitung (~125k articles), and NZZ am Sonntag (~9k articles) and the recruited 23 professionals, we generated a high-quality dataset consisting of a stance label for a given question and articles and an

emotion label for a given article text on paragraph level and full-article level. After post processing methods, we created the dataset called 'CHeeSE': The German Dataset of Swiss (CH) News Articles for Stance and Emotion Detection.

The dataset has 91 debate questions and 670k Swiss news articles (Blick, NZZ, NZZaS) with 3'693 annotated question-article pairs (Stance, paragraph emotion, article emotion). Specifically, the dataset comprises a total of 1,970 different articles paired with up to six questions (~1.87 on average) from a list of 91 questions that we manually compiled to ensure a variety of debate topics. We released the dataset for the research community and performed target-specific stance detection as a supervised task using the CHeeSE dataset and provide baseline models as a reference in future research on stance detection in German news articles.

Research Prototypes and Use Case

We built an emotion viewer tool and stance viewer tool using the dataset we created (see Figure 4). The emotion viewer, for a given text from a news article, will predict the emotions that the text conveys. The stance viewer displays articles grouped by their stance. For example, given the topic "exit nuclear energy", the user will obtain an overview of news articles that are «in favor», «against», or «discussing» the topic. We further explored the usage of expressed emotions for recommender systems where the aim is to incorporate the user's and article's expressed emotions into a model to generate accurate news article recommendations. Currently, we are planning to conduct live user study with our prototypes. ■

- We built an annotation tool that enables annotators to provide a stance label for a given question-article-pair and an emotion label for a given article text on paragraph level and full-article level.
- We created the CHeeSE dataset: The German Dataset of Swiss (CH) News Articles for Stance and Emotion Detection with 91 debate questions, 670k Swiss news articles (Blick, NZZ, NZZaS) and 3'693 annotated question-article pairs.
- We released the dataset for the research community and performed offline experiments on the CHeeSe dataset for both detection tasks.
- Built two research prototypes: emotion viewer tool and our stance viewer tool.
- Our experiments demonstrate that usage of expressed emotions from articles boost the recommender systems performance.

Intelligent Content Creation Tools for Augmented and Virtual Reality

To unleash the full potential of extended reality (XR), high-quality interactive experiences are crucial. This project, in collaboration with the Game Technology Center and Disney Research Studios, aims to develop tools to create intelligent, story-driven characters, making the XR world a more compelling human experience.

The augmented and virtual reality (XR) revolution is approaching. Companies across the world are working toward XR technology that provides a seamless merger of real and virtual content to intertwine contextualized digital information with intelligent interfaces that connect citizens closer to the world around them. Such platforms will be driven by the confluence of sophisticated technologies reaching maturity. While a tremendous effort is already being invested in core enabling technologies to create and deliver XR worlds, almost no attention has been directed at populating them. In this project, we address this limitation – our goal is to develop tools for intelligent story-driven digital characters that inhabit XR environments, transforming XR into a compelling human experience.

Towards this goal, the Media Technology Center (MTC) has been collaborating with the Game Technology Center (GTC) and DisneyResearch|Studios (DRS). The final goal of this collaboration is to have autonomous virtual characters that can navigate and interact with the real environment. Developing such characters relies on key components such as scene understanding, animation synthesis, story technology and character intelligence. In the course of this project, we have developed two use-case prototypes so far: AR Museum Guide, and AR Explorer.

Figure 5: AR explorer character navigating the real world.

Figure 6: AR characters interacting with the real environment.

AR Museum Guide

In this experience a virtual AR character leads the user through a real art exhibition, while interacting with the user and the surroundings. After the user has been greeted, the virtual character takes the user on a tour through the museum, making stops at paintings, explaining their background and pointing out details. In order to enable the character to interact with the user and with real-life paintings, we have made technical progress in the areas of scene understanding and animated character behavior authoring. Our prototype combines these advances with natural language interactions, smart navigation and animation synthesis in order to create a compelling AR experience.

Figure 7: AR museum guide in an exhibition.

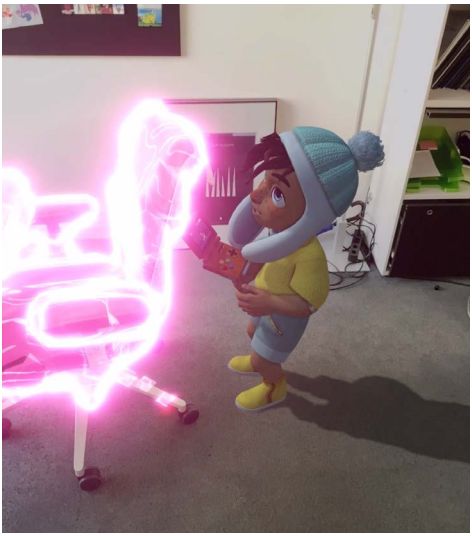

Figure 8: AR explorer character discovering objects.

AR Explorer

In the AR Museum Guide scenario, the scope of the character-user interaction is generally limited by the behaviors authored by the 'curator'. Going beyond such constraints, we have developed a second use-case prototype in which an animated AR character freely explores the real world and finds interesting objects in the real environment with the user's help. The character observes the surroundings, attempts to understand the objects in the scene, and interact with them in a playful manner. An important challenge in actualizing such an interaction between the character and the scene is the variability and unpredictability of the real world. As the real environment is subject to constant change and cannot be known in advance, it is essential to develop a system that can enable the characters to dynamically acquire knowledge about the world, and use this set of knowledge to plan their behaviors. The higher level of technical challenges present in this prototype requires us to go beyond the technologies used in the museum guide prototype.

In order to actualize this prototype, we have advanced our scene understanding technologies, allowing for a more precise scene object segmentation. Furthermore, the system we developed enables the character to predict scene affordances, i.e. the interaction capabilities offered by the objects in the scene. The ability to predict affordances allows the virtual XR characters to believably interact with the real environment through the animation system. In addition, our application is able to benefit from Large-Language Models to enable free-form conversations between the character and the user, next to conversations that are bounded by pre-scripted stories. All of these components are bound together thanks to our improved story system which enhances the overall quality of the AR experience. ■

- A prototype of an intelligent virtual museum guide character was developed for the AR Museum Guide.
- The AR Museum Guide prototype was showcased at the Intelligent Content Creation Tools for AR/VR Workshop '22 in Ascona, Switzerland.
- A second prototype, the AR Explorer, was created with advanced scene understanding and interaction capabilities.
- Collaborations were established with Disney Research Studios, Game Technology Center, and partners in South Korea.
- A total of six Bachelor theses, one Semester thesis, and six Master theses were supervised together with GTC and DRS.

Education

We are dedicated to introducing the new generation of computer science students to the exciting possibilities in media technology alongside our research projects.

In recent years, the rapid progress in machine learning and artificial intelligence (AI) has sparked significant enthusiasm among ETH students. With their immense potential spanning numerous disciplines, we are dedicated to acquainting students with the distinct complexities and prospects inherent in data-driven solutions within the news and media sector.

Seminar on Media Innovation
For the fourth consecutive year, we have organized a student seminar aimed at introducing Bachelor students to the applications of machine learning and AI in media technology. Covering topics such as language understanding, fake news detection, generative models for music and audio, and image synthesis and recognition, the seminar offers students a comprehensive understanding of how data-driven algorithms can drive innovation and create captivating content.

Student projects
In 2022, we supervised a total of five student projects, offering hands-on opportunities for students to develop new algorithms, experiment with state-of-the-art code, and validate their ideas. These projects covered diverse topics within our project portfolio, enabling students to gain practical experience in applying machine learning and AI techniques to real-world media challenges. ■

All student projects can be accessed on our website.

Outreach

The MTC has been actively engaging in both physical and virtual events to showcase the outcomes of our research projects, as part of our ongoing outreach efforts.

Events

21 March	**Aesthetic Assessment of Image and Video Content @ TX Group, Zurich** – Brownbag lunch talk with project insights and showcase of first results.
28 March	**Multilingual Abstractive Article Summarization @ TX Group, Zurich** – Brownbag lunch talk with insights about the project.
4 April	**MTC Showcase @ Ringier, Zurich** – Presenting an overview of the MTC's projects, together with a demo session.
21 April	**MTC @ SACM Conference, Zurich** – Presenting an overview of the MTC to the Swiss communication and media research community.
2 May	**EqualVoice Summit, Zurich** – Presentation of the EqualVoice-Factor evaluation and project with Ringier.
19-23 June	**Intelligent Content Creation Tools for AR/VR @ Congressi Stefano Franscini, Monte Verità, Ascona** – Workshop held as part of a larger research collaboration between Switzerland and South Korea funded by Innosuisse and KIAT.
October November	**ZHdK Module** – Nine design students at Züricher Hochschule der Künste created prototypes and demos for MTC research projects in the Media Innovation Lab module.
October November	**Exchange & Coworking @ MTC** – Milena Djordjevic (TX Group) works from the MTC for a two week exchange.
21 December	**Ringier Journalism School Visit @ MTC offices, Zurich** – Presenting an overview of MTC and demos showcase.

Scientific Publications

RecSys Challenge 2022: Fashion Purchase Prediction
Nick Landia, Frederick Cheung, Donna North, Saikishore Kalloori, Abhishek Srivastava and Bruce Ferwerda
16th ACM Conference on Recommender Systems (RecSys '22), Seattle, WA, USA

RecSys Challenge 2022 Dataset: Dressipi 1M Fashion Sessions
Nick Landia, Rachael Mcalister, Donna North, Saikishore Kalloori, Abhishek Srivastava and Bruce Ferwerda
ACM Recommender Systems Challenge Workshop (RecSys Challenge 2022), Seattle, WA

Fast content placement and alignment in 3d scenes
Robin Wiethüchter, Saikishore Kalloori and Ribin Chalumattu
Proceedings of the Thirty-Fifth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2022, Gainesville, FL

Cross-silo federated learning based decision trees
Saikishore Kalloori and Severin Klingler
37th ACM/SIGAPP Symposium on Applied Computing (SAC 2022), New York, NY

Personalized Information Retrieval for Touristic Attractions in Augmented Reality
Felix Yang, Saikishore Kalloori, Ribin Chalumattu and Markus Gross
15th ACM International Conference on Web Search and Data Mining (WSDM '22), New York, NY

Understanding Aesthetics with Language: A Photo Critique Dataset for Aesthetic Assessment
Daniel Vera Nieto, Luigi Celona and Clara Fernandez-Labrador
36th Conference on Neural Information Processing Systems (NeurIPS Datasets and Benchmarks 2022), New Orleans, LA, USA

All of MTC's scientific publications can be accessed on our website.

ETH Zürich
Media Technology Center
STF
Stampfenbachstrasse 114
8092 Zürich

Contact:
+41 44 832 82 86
fabiopo@ethz.ch
www.mtc.ethz.ch

The Media Technology Center is supported by Ringier, TX Group, SRG SSR, NZZ, Verband Schweizer Medien, viscom and the ETH Zurich Foundation.