DATA TECHNOLOGY SEMINAR
AN EBU EVENT

PROGRAMME
TUESDAY 12 MARCH 2024

Timings in CET

09:00 – 09:55   **Opening** Antonio Arcidiacono (EBU)

INNOVATION SHOWCASE: NEXT GENERATION OF MEDIA SOLUTION
Moderated by Hans Hoffmann (EBU)

09:05 – 09:15   **Data, AI, and Large Language Models – how the EBU can help you innovate**

We present our News Hub, and associated AI workflows for summarisation, entity extraction, and our latest experiments with Large Language Models and how these tools can help the journalists in the newsrooms.

Sébastien Noir (EBU) Active since 2012 in the Broadcasting industry, first at RTS the french speaking Swiss National Broadcaster as a software developer, Sébastien Noir evolved to lead the development of multiple digital products and mobile applications. He then became the Product Manager of the VOD Platform PlaySRG for Switzerland, coordinating development teams, and delivering multilingual products for the different linguistic region.
In 2017, he joined the European Broadcasting Union to work as Product Owner for PEACH, the Personalisation and Recommendation System developed by Broadcasters for Broadcasters.
He is now head of Software engineering in EBU Technology and Innovation Department, coordinating teams and developments effort of innovative Services like PEACH, EuroVOX, and the EBU News Pilot making use of artificial intelligence.

Overview of PEACH, transparent AI data platform, enabler for creating personalised experiences for citizens, by providing content recommendations and analysing user behaviour

Dmytro Petruk (EBU) Product Owner @ PEACH

09:15 – 09:20   **ThoughWorks** Gabriel Gavasso (Thoughworks)

09:20 – 09:25   **Cloud-native video production and collaboration** Kim Eikeland (mimir)

09:25 – 09:30   **Empowering Innovation: Unlocking Agility with No-Code for Future-Ready Media Workflows** Dennis Laupman (PluxBox)

09:30 – 09:35   **Ratings Artist: an ML-based rating prediction framework to incorporate human expertise** Jens Coster (MediaGenix)

09:35 – 09:40   **Enhancing journalistic workflows with Wolftech** Sergej Stoppel (Wolftech)

09:40 – 09:45   **True Confidence Language Model Process** Johan Classon (Conversy.ai)

09:45 – 09:50   **Personalized Media Recommendations driven by Knowledge Graphs & Large Language Models** Alexander Schulze (Innotrade)

09:50 – 09:55   **Product impulse: No Nonsense AI. Confidence scores and LLM routing for media applications** Jonas Petersen (Media Press / K2 AI)

09:45 – 10:15   ***Break and demos***

DATA TECHNOLOGY
SEMINAR
AN EBU EVENT

SESSION 1: MEDIA DATA GOVERNANCE – PAVING THE WAY FOR INNOVATION

Moderated by Mike Matton (VRT)
https://www.linkedin.com/in/mikematton/

10:15 – 10:45   **Data Strategy for the BBC Archives**

An overview the BBC strategy to unify our archive catalogues, enrich with machine learning and semantic tagging and build a flexible architecture.. I'll talk about our challenges, priorities and plan to get the most out of our content and data.

Richard Jolly (BBC) In my 18 years at the BBC I've helped build several of the core data systems. I enjoy how the fundamentals of good data practices meet the realities of legacy architectures and pressing business needs. I've recently found my spiritual home in the BBC Archive Technology & Services, where we are working to transform our data systems to get the best from our content. https://www.linkedin.com/in/richard-jolly-94041883/

10:45 – 11:15   **From Data to AI: Harmonizing innovation with mature governance**

Samuel Profumo (RTBF) Currently Chief Data Officer of Radio Television Belge (RTBF), public service broadcasting organization in Belgium. RTBF operates five television channels – La Une, La Deux, La Trois, Arte Belgique and PureVision together with a number of radio channels, La Première, RTBF International, VivaCité, Musiq3, Classic 21, and PureFM. My role encompasses data governance, data strategy and data protection. Previously, I used to work for Groupe Figaro, the 1st french media destination in France on Internet. After graduating from Institut Mines-Télécom Business School with a Master in strategic management, I started my career in print and digital advertising where I led operations, revenue management and digital strategy. I've been in charge of digital tools and analytics transformation, and contributed to the launch of different key initiatives, such as La Place Media, the first Premium Publishers Adexchange in Europe. During the last 4 years, I started and led a 40 people multidisciplinary data team (data architects, analysts, scientists…) dedicated to leveraging data value by creating and implementing new data use cases for the three main activities of Groupe Figaro: Publishing (Le Figaro, La Chaîne Météo, L'internaute, Comment Ça Marche...), Sales (Viadeo, Figaro Store, Cadremploi, Ticketac, Marco Vasco, Les Maisons du Voyage, Explorimmo...) and Advertising (MEDIA.figaro, CCM Performance, FigaroClassifieds, Zebestof Trade Desk.. https://www.linkedin.com/in/samuel-profumo-62ba25

11:15 – 11:45   *Break and demos*

SESSION 2: THE AI REVOLUTION IN MEDIA
Moderated by Hanna Lukashevich (Fraunhofer IDMT)

Hanna Lukashevich is head of the Semantic Music Technologies (SMT) research group at Fraunhofer Institute for Digital Media Technology IDMT in Ilmenau, Germany. Among her research interests are audio signal processing and machine learning. As head of the SMT group she manages a number of R&D projects, both for companies in the industrial and digital media sector as well as national and international public projects. **https://www.linkedin.com/in/hannalukashevich**

11:45 – 12:15   **Enhancing data accessibility and generation: Leveraging Data Hubs, LLMs, Graph Relations, and Natural Language Queries for Metadata**

This presentation showcases the integration of Language Models (LLMs), Graph Relations, and Natural Language Queries to innovate in the Media and Entertainment industry by enhancing metadata accessibility and generation. By aggregating diverse data sources into a unified data lake and utilizing AI to enrich metadata quality, the approach offers a data hub for analytics and AI analysis. The creation of a graph model, where each data point is interconnected, facilitates natural language interactions and the generation of new, relevant metadata. Emphasizing the use of standard, Retrieval Augmented Generation (RAG)-enriched, and industry-specific fine-tuned LLMs, the presentation demonstrates how these technologies enable intuitive, efficient content production and insights generation. This innovative combination breaks down data silos and optimizes the media supply chain, making a significant leap towards a more connected and productive media and entertainment workflow.

Robert Raver is a Sr. Media and Entertainment Specialist Solutions Architect and Media Supply Chain tech lead, focusing on Media Content Creation and Media Supply Chain's at AWS. Robert designs innovative cloud-based architectures that optimize media supply chains, solving industry challenges related to media processing, rights, title,

---

QC, and metadata management. Robert has over 20 years of experience in various industry roles, including leadership, architecture, and engineering. https://www.linkedin.com/in/rraver/

Roland Duboué (Amazon/AWS) After starting his career as a musician then as a sound designer in for movie pictures, Roland developed his activity around Media Asset Management in major French channels (TF1, France Télévisions). He now works at AWS as a Solutions Architect specializing in Media Supply Chain for European customers. **https://www.linkedin.com/in/roland-duboue-84590b138/**

12:15 – 12:45   **Media's AI Frontier**

Google Cloud's perspective on the future of media with AI, focusing on content production, personalisation of audience experiences and archives.

Maria Alonso Garcia is a Customer Engineer at Google Cloud focusing on media solutions. We work closely with media companies and discuss new technological advances. https://www.linkedin.com/in/mariaalonso2705/

12:45 – 14:30   *Buffet Lunch in the foyer and demos*

SESSION 3: DEMYSTIFYING AI – EXPLORING TRANSPARENCY AND REGULATION

Moderated by Alexandre Rouxel (EBU)

Alexandre Rouxel (EBU) is Senior Project Manager for Data and AI at the EBU, where he manages the AI and metadata community, AIM, and coordinates collaborative working groups on metadata, AI and cloud computing. At T&I, he leads the development of the EBU AI-HUB, a cloud-hosted platform to showcase, evaluate and share AI applications for media designed by and for Members, including: facial recognition for TV programmes, AI to enrich radio programmes and a fake news analyser.

14:30 – 15:00   **Peeking into the black-box of AI**

Machine Learning (ML) has revolutionized industries, driving innovations in healthcare, finance, and beyond. Its capacity to analyze complex datasets and predict outcomes has made it invaluable. However, ML's effectiveness is tempered by the "black-box" problem, where the decision-making process of algorithms remains opaque, undermining trust and accountability. This talk will delve into Explainable Artificial Intelligence (XAI), offering insights into how XAI seeks to make ML models more transparent and understandable, thus bridging the gap between technological capability and ethical responsibility.

Alan Perotti, PhD is a computer scientist from Turin, Italy. He works as a researcher @ CENTAI Institute. His fundamental research focuses on explainable AI and neuro-symbolic integration. He is part of industrial projects on applied ML for finance and a Horizon Europe project on ML for healthcare. https://www.linkedin.com/in/alan-perotti-phd-146b4037/

15:00 – 15:30   **Benchmarking and survey of explanation methods for black box models**

The talk will present an overview of evaluation measures for the most common explainability techniques, ranging from saliency maps to counterfactual explanations, with a conclusive case study on health.

Riccardo Guidotti is an Assistant Professor (RTD-B) at the Department of Computer Science at the University of Pisa and a member of the Knowledge Discovery and Data Mining Laboratory (KDDLab), a research group in collaboration with ISTI-CNR of Pisa and with Scuola Normale Superiore. Riccardo Guidotti graduated with honors in Computer Science (Master's and Bachelor's degrees in 2013 and 2010, respectively) from the University of Pisa, where he also earned his Ph.D. in Computer Science with a thesis on "Personal Data Analytics". He won the IBM fellowship program and worked at IBM Research Dublin, Ireland, in 2015, the DSAA New Generation Data Scientist Award in 2018, and the Marco Somalvico Award for Artificial Intelligence in 2021. His research interests include explainable artificial intelligence, interpretable machine learning, personal data mining, clustering, and the analysis of transactional data and time series.
https://www.linkedin.com/in/riccardo-guidotti-b89b3280/

| 15:30 – 16:00 | **The emerging AI regulation** |

This presentation will take us through the different shapes of emerging AI regulation across the world and assess how this may impact the European media industry.

François Lavoir (EBU is an EU policy expert with 10 years of experience in media and digital regulation. He leads the European Broadcasting Union's policy activities in Brussels in the fields of artificial intelligence, data protection and digital competition. In his role, he closely followed the European Union's negotiations on the Artificial Intelligence Act and the potential impact of this new law on public service media.

| *16:00 – 16:30* | *Break and demos* |

SESSION 4: COUNTERING DISINFORMATION – INNOVATIVE APPROACHES

Moderated by Paolo Casagranda (RAI)

Paolo Casagranda is a researcher at Rai R&D (CRITS), with a passionate interest for technology and its impact on people. He obtained a PhD in Computer Science at the University of Torino, investigating recommender systems applied to audio and radio. His work is focused on radio personalization, automatic metadata extraction and the application of AI tools to audio content.

| 16:30 – 17:00 | **Addressing fake news with reliability** |

Detecting fake content is a sensitive task. On the one hand, we want to be able to address as many cases of disinformation as possible. Yet, on the other hand, wrongfully accusing some content of being fake can be harmful to the wrongfully-accused authors. Ultimately, those false positives can erode the public trust in fact-checking tools, and hinder the impact of debunking operations. In this talk, I will present the a contrario paradigm, which enables making automatic detections from the results of fact-checking tools, by mathematically analysing the methods to threshold on the tolerated rate of false positives. A contrario analysis can thus ground detections by showing how rare such detections should be on real content.

Quentin Bammey received the Ph.D. degree in applied mathematics from ENS Paris-Saclay, Université Paris-Saclay, France, in 2021. He is currently a researcher in multimedia forensics, image processing, computer vision and machine learning at the École Normale Supérieure (ENS) Paris-Saclay, Université Paris-Saclay, Centre Borelli, CNRS, France.

Dr. Bammey is mainly credited with the invention of positional learning (Q. Bammey, R. G. von Gioi and J. -M. Morel, ``An Adaptive Neural Network for Unsupervised Mosaic Consistency Analysis in Image Forensics,'', CVPR 2020. As part of the SaclAI-School project, he is also participating to the coordination and development of the ``BrevetAI'' pedagogical platform, to offer a learning-by-doing training on artificial intelligence. He also organizes the IPOL MLBriefs workshop (https://mlbriefs.com/) to promote reproducible research and facilitate the reproducibility of existing methods. His research interests include multimedia forensics, image and video processing, computer vision, and machine learning. **https://www.linkedin.com/in/quentin-bammey-802290173**

| 17:00 – 17:30 | **Unveiling AI solutions for disinformation combat: Insights from the IDMO project** |

Lorenzo Canale (RAI)

| 17:30 – 18:00 | **Digital Traces: Verification of audio-visual content** |

This presentation will delve into selected aspects of disinformation related to audio-visual media. It will categorize various types of disinformation and discuss general strategies for countering disinformation, focusing on the differentiation between the falsification of claims about content and the authentication of the content itself. Using the audio domain as example, it will present current analysis and detection techniques, and outline the key challenges in the field. Finally, it will demonstrate how the presented aspects are addressed in current collaborative projects like SpeechTrust+, VERA.AI, News-Polygraph, and GEISST.

Patrick Aichroth worked as a freelance software developer and IT trainer before becoming a research associate at Fraunhofer IDMT in 2003. Since 2006, he has been head of the Media Distribution and Security Group at Fraunhofer IDMT, which focuses on the development of technologies for audio manipulation and synthesis detection, audio provenance analysis, media security, privacy enhancing technologies and trustworthy AI. **www.linkedin.com/in/patrick-aichroth-a131a9**

Hanna Lukashevich (Fraunhofer IDMT)

| *End of Day 1* |

SESSION 5: UNLOCKING INNOVATION – SELF-DESIGNED AI TOOLS FOR ADVANCED MEDIA INTELLIGENCE (ROOM MONTREUX)

Moderated by

09:00 – 09:30   **Advancing media intelligence: The Face Management Framework (FMF) for automated annotation and diversity analysis for public service media**

The Face Management Framework (FMF) is the system Rai developed within AI4Media, an EU H2020 funded project, for automated annotation of TV personalities in video streams and archives. This talk will describe how we developed the Face Management Framework and is intended to be a starting point for deepening a discussion about tools and methodologies to support AI-based diversity analysis of A/V content.

Maurizio Montagnuolo is a Senior Research Engineer working at the R&D Department of the Italian public broadcaster RAI. He graduated in Telecommunications Engineering and holds a Ph.D. in "Business and Management". His interests are mostly addressed in the context of multimedia data mining and artificial intelligence, fields in which he counts several publications in international journals and conferences. He has been working on several EC funded projects in the field of digital archiving, automated metadata extraction and cloud technologies. https://it.linkedin.com/in/maurizio-montagnuolo-17339442

09:30 – 10:00   **Diversity knowledge graphs**

Using linked open data to assess the diversity of the programme portfolio

Jo Kent is a Data Architect in BBC Research and Development. She has worked with BBC data in one form or another for over 15 years now, with a particular focus on semantic linked data and knowledge graphs. She has presented papers and talks at semantic linked data and taxonomy conferences nationally and internationally and is always happy to talk and learn more about data, graphs, ontologies and taxonomies. https://www.linkedin.com/in/kentjo/

10:00 – 10:30   **Enabling media professionals to create their own datasets and AI tools**

Georg Thallinger studied Telematics at Graz University of Technology and has been working at JOANNEUM RESEARCH since 1992. He heads the "Smart Media Solutions" group within the "Intelligent Vision Applications" research group of DIGITAL. His research focuses on methods in the area of content-based analysis and search of audiovisual media for applications in media production and monitoring, audiovisual cultural heritage as well as mobility and security. He is the coordinator of FAIRmedia and DIDYMOS-XR and has led a number of national and international projects (e.g. DIAMANT, SALERO, FascinatE, SIMMARC, SIMPLE, TailoredMedia).

Christoph Bauer (ORF) was born in 1960 in Vienna/Austria, studied at Vienna's University of Economics and has several other qualifications like cantor, pianist, organist, choir-conductor, composer, IT-developer, theologian, etc.; when starting to quit fooling around, he joined ORF in 1981; his main tasks since then (excerpts): He acted as Project Officer ORF for several EC/IST/ICT/H2020/FAA-Projects and is the senior specialist for preservation, digitization and restoration in the ORF archive department. In addition, he is acting as system-administrator for Archive-Systems and AV-Digitization, workflow-developer and AI-mining-specialist (audio&video). Christoph was chairman of the SNML-TNG Management Board (2011-2013), vice-chair of maa (Media-Archives-Austria Association) (2012-2016), member of the ARD K-ARL Expert group for Video-Mining, NKE for the EU-Project "Empowering Society" and lecturer at the University of Vienna. He is the current general-secretary of maa (Media-Archives-Austria Association), member of the ARD medas Expert group Mining (AI) and member of the Digitization & Migration Commission of FIAT/IFTA.

Gernot Rottermanner (FHSTP)

DATA TECHNOLOGY
SEMINAR
AN EBU EVENT

SESSION 6: INTEGRATING AI TOOLS INTO YOUR ORGANIZATION (ROOM GENEVA)
Moderated by Hanna Lukashevich (Fraunhofer IDMT)

09:00 – 09:30   YleGPT – how to get everybody on board with AI

In his presentation, Jyri will share insights into how Yle has successfully navigated the AI revolution. Attendees can expect a compelling narrative that not only highlights Yle's journey through the AI boom but also offers practical strategies for implementing AI solutions in editorial work.

Jyri Kivimäki (Yle) As the Editorial AI Solutions and Innovation Coach at Yle, Finland's largest media company, Jyri Kivimäki has dedicated almost a decade to pioneering the integration of artificial intelligence within journalism. Working within the Yle News Lab, part of the News and Current Affairs development unit, Jyri is at the forefront of ensuring that technological advancements are not just kept pace with but are harnessed to drive forward-thinking journalism. In his role, he has been instrumental in ensuring that no team member is left behind in the rapidly accelerating tech landscape, fostering an environment of continuous learning and adaptation. https://www.linkedin.com/in/kivimaenjyri

09:30 – 10:00   AI supported helpdesk

Presentation will cover AI helpdesk application development process and explanation

Ruhi TAŞ is Deputy Director of Information Technologies Department at Turkish Radio Television Co. He received the Bachelor's and Master's degree from the Hacettepe University, Electrical and Electronics Engineering Department Ankara - Turkiye in 2011. Also received PhD degree from the Ankara University of Computer Engineering Department, Turkey. He studied as a software and R&D engineer with extensive experience and management skills for several projects more than 27 years. His research interests include RF radio systems (FM, DAB, Antennas, remote monitoring e), HBBTV (IBB) systems, blockchain technology, and AI systems. https://www.linkedin.com/in/dr-ruhi-t-8932411ba/

10:00 – 10:30   SaaS + AI + Journalism = Gold

SaaS and cloud are the true enablers of AI tools, and only when their advantages are combined can they

André Torsvik has 20 years of experience in television production and the media tech industry, working for small startups and large multinationals, doing everything from editing through production to tech marketing. He now works for Mimir, a scale-up providing cloud-native AI-friendly asset management as a service. https://www.linkedin.com/in/andretorsvik/

10:30 – 11:00   Break and demos

SESSION 7: REVOLUTIONIZING MEDIA WITH AI AND GENERATIVE AI (ROOM MONTREUX)
Moderated by Paolo Casagranda (RAI)

11:00 – 11:30   Unlocking treasures: Enhancing discoverability of French audiovisual heritage with speech-to-text

France's national audiovisual institute (INA) mission is to collect, save, digitize, preserve and promote French television and radio archives. Guardian of audiovisual heritage, the INA provides unique experience and know-how in the structuring and promotion of its archives, in an approach focused on users and clients from around the world. With a mission to collect, digitize, store and promote the content of 184 television and radio channels, the INA takes up the challenge of exhaustively describing this vast archive. Currently, only a part of the archive is detailed through external metadata imports or manual processing, requiring a transformative approach. In line with our commitment, INA plans to harness the power of artificial intelligence (AI) to automate and accelerate archiving processes. At the heart of most AI treatments, we identify the speech-to-text process as one of the pillars of most treatments. With our ambition to convert 28 million hours of audiovisual content into text, we aim to unlock a wealth of metadata generating titles, summaries, semantic segmentation and predictions of named entities and thematic categories. The driving force behind this ambitious initiative is to minimize human efforts and time invested in manual processing. Choosing the right tool is essential to the success of this initiative. The presentation provides an overview of the challenges ahead, our transformation ambitions and critical considerations in selecting a tool, focusing in particular on processing speed and evaluation metrics. Our ultimate goal is not only efficiency but also enrichment of accessibility and understanding of the cultural and historical audiovisual narrative of France. The presentation will offer insight into the delicate balance we seek between automation and human intervention.

Eleni Kogkitsidou is a data project manager at INA's data and media management service, and she holds a PhD in Natural Language Processing from the Grenoble Alpes University. In the previous years, she worked in private and public companies as a researcher on issues related to textual data processing and analysis, as well as a lecturer in computational linguistics at several French universities. In addition to her role as a data project manager, Eleni also serves as the Product Owner of NLP projects, overseeing the end-to-end execution of various NLP initiatives within the organization. Her expertise in NLP, honed through years of research and practical experience, is instrumental in guiding these projects to success. Furthermore, Eleni takes charge of the coordination of process management, ensuring that workflows are effectively implemented, and that all stages of data and media management align with the organization's goals and objectives. Her multifaceted role combines academic knowledge and practical skills to drive innovation and efficiency in the field of data and media management.
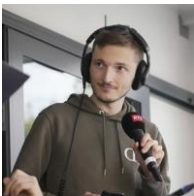

11:30 – 12:00   The EBU AI Hub: A deep dive into the radio programme enrichment application developed with RadioFrance

Projet in collaboration with the EBU to chapterize radio podcasts by identifying the key host questions

Ivan Thomas (RadioFrance) I love listening podcasts and learning new Natural Language Processing technics.

Alexandre Rouxel (EBU)

Marin Piguet is a student in digital humanities at the Swiss Federal Institute of Technology in Lausanne (EPFL) and works as an intern at the EBU on his Master's thesis on podcasts automatic enhancement with AI and data science tools. He also works as a radio host for the classical music productions of the Radio Television Suisse. These two activities give him a critical and comprehensive overview on the radio media of yesterday and tomorrow.

12:00 – 12:30   Large-scale deployment of GenAI for metadata enrichment – a tale of yield and caution

Aleksander Obuchowski (Media Press)

SESSION 8:   DATA MASTERY IN MEDIA – FROM IN-HOUSE ANALYTICS TO AUTOMATED CONTENT CREATION (ROOM GENEVA)

Moderated by Tatjana Mladenovic (BBC)

Lead Data Manager at the BBC Archives Technology & Services Data Platform team, working on content metadata. She has been working on metadata for media in engineering, data analytics and management roles for almost 20 years. Tatjana's main interests are automation of metadata processes, data modelling and data quality.

11:00 – 11:30   Taking back control of usage data: in-house tracking

A few years ago, Radio France decided to set up an in-house tracking system to replace the 3d parties' trackers. The aim is to regain ownership of the data, have all listening data certified, and above all to be able to offer innovation based on this usage data. This conference will look back at the entire data cycle, from collection to analysis, including the questions raised and the tools used.   First, we'll take a look at what Radio France distributes (live & podcast, on different media), and what solutions have been implemented to collect this data. Next, we'll look at how we enrich this data from internal metadata and the choice of our Data Warehouse. Finally, we'll take a look at all the uses : analytics, monitoring, recommendation, product needs, etc.
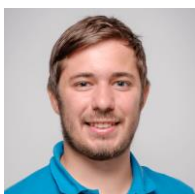
Bastien LUNETEAU (Radio France) Lead Data Engineer for Radio France, in charge of in-house tracking, search engine and data science industralization. https://www.linkedin.com/in/bastienluneteau/

11:30 – 12:00   **Yle's first Year in review: Wrapping personal data and insights into a visual mobile experience**

A demonstration of the feature + peak under the hood

Riikka Lätti (Yle) Riikka oversees and helps with the development of personalization technologies in 3 teams that focus on recommendation, interaction and sign-in services. In her day-to-day work, she works somewhere between strategic and operational level, enabling communication and decision-making between dev teams and internal stakeholders. For this project (Yle's year in review), Riikka worked as a project manager.

12:00 – 12:30   **AI usage of NHK archives**

NHK has built an ultra-large-capacity storage system to manage approximately 55PB of file materials for past programs broadcasted and programs to be produced in the future. We will introduce the current state of contents management at NHK Archives and our efforts to shift to the cloud. In addition, as expectations for digital archives increase, it is expected that the value of the large amount of video and audio materials stored in the NHK Archives will be maximized. Therefore, we will also report on our efforts to improve the sophistication of metadata using image recognition AI and voice recognition AI to improve the search performance for the content that users are looking for.

Masaharu Ito is an engineer at NHK, working on the development and construction of archive systems. Masaharu is actively working to introduce the latest technology into archive systems, such as building an automatic anonymization system for news manuscripts using NLP and an automatic metadata generation system for sports videos. Masaharu likes sushi, but is not good at fishing.
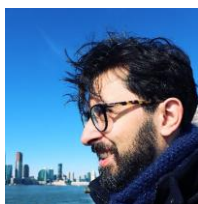
12:30 – 14:30   *Buffet Lunch in the foyer and demos*

SESSION 9: REDEFINING RECOMMENDATIONS IN PSM – EMOTION, VALUES, AND AUDIENCE INSIGHTS
Moderated by Alexandre Rouxel (EBU)

14:30 – 15:00   **Recommenders with values: Developing recommendation engines in a public service organisation**

Alessandro Piscopo is a Lead Data Scientist at BBC, Product Group. His team develops recommendation engines across the organisation and has deployed recommenders on products such as BBC Sounds, BBC World Service, and BBC iPlayer. Alessandro's research interests lie at the intersection of information retrieval, collaborative knowledge engineering, data quality, and responsible AI. He received his PhD from the University of Southampton in 2019. www.linkedin.com/in/alessandro-piscopo-ds

15:00 – 15:30   **What are you in the mood for? The potential of emotion-based recommendations on streaming platforms**

Human behaviour is very much driven by emotion. So it follows that emotions also influence our media consumption – such as in how we decide what we want to watch on video streaming platforms. Understanding the emotional effect of content supports improvements in both content recommendation and engagement as well as content development. The presentation gives an introduction to how metadata and AI could be used to recommend content on a video streaming platform, based on the mood and feel it gives to the audience. It will also suggest ideas for the future in this field of emotional recommendations, using AI with humans in the loop. In addition to its opportunities and benefits, an emotion-based recommendation also raises considerations about ethics and risks.

Elina Yli-Ojanperä is a Publishing executive at the Finnish public media service company Yle. She specializes in working with digital content and platforms to provide an objective view from the customer's perspective. Her expertise includes supporting digital content development, creating publishing plans, and understanding digital and streaming platform features and data analysis. She has a strong interest in series, digitalization, emotions, and the integration of artificial intelligence with human involvement. She follows the emotional impact from storytelling to publishing and offering content. Contact: elina.yli-ojanpera@yle.fi

15:30 – 16:00   **Introducing Kaleidoscope**

Constructing a tool that visualises and quantifies viewing in the UK television landscape.

Connor Laughey is a Senior Data Scientist at the BBC in the Chief Customer Officer group focussing on the iPlayer and Sounds streaming services. He is currently developing the Kaleidoscope app. He has also worked on evaluating the first personalised one to one marketing campaign on iPlayer and analysed the effects of sporting events and user behaviour on the Sounds app. Prior to joining the BBC, Connor held a data science position at the UK's Department for Education and attained a Master's degree in Data Science from the University of Manchester. https://www.linkedin.com/in/connor-laughey-1b989812b/

John Davies is a Principal Data Scientist at the BBC in the Chief Customer Officer Group, where he focuses on the iPlayer and Sounds streaming services. Prior to this he worked in the data science team at the food delivery company Deliveroo on customer conversion and experience. He also worked in the data science team at Nesta, the UK's innovation foundation, applying data science to issues in digital and creative industries. Projects included quantifying the adoption of AI in creative industries and evaluating networking at tech conferences. @johnardavies, https://www.linkedin.com/in/john-davies-31858217/

16:00 – 16:30   *Break and demos*

SESSION 10: THE ROLE OF STANDARDIZED DATA AND ONTOLOGIES IN ENTERPRISE TRANSFORMATION
Moderated by Tatjana Mladenovic (BBC)

16:30 – 17:00   **Becoming a data-driven media enterprise with EBUCorePlus**

This talk presents how EBUCorePlus can describe the big picture for data in media enterprises, which projects have used it recently and why media enterprises need it to become data driven.

Jürgen Grupp (SWR) is a member of the board of EBU's Artificial Intelligence and Metadata group (AIM). He is leading the development of EBUCorePlus in the Metadata Modelling initiative. As information architect at SWR, he's trying to promote data literacy throughout the organisation. https://www.linkedin.com/in/juergen-engelbert-grupp/

17:00 – 17:30   **Leading the evolution: The impact of the Rai Ontology on Enterprise Information Architecture**

The transition from traditional broadcaster to media company is only a part of the ongoing digital transformation. RAI uses Enterprise Architecture to identify Information Architecture that guarantees full interoperability among business domains. Semantic alignment is addressed by an enterprise ontology based on EBUCorePlus

Marco Riccobene (RAI) Has gained considerable experience in the IT field on various subjects from Application Development to IT Strategy, passing through IT Service Management, Project Management and Business Analysis. He has been dealing with Enterprise Architecture for a couple of years. Holds a master in IT Governance and Compliance and his interests are mostly aimed at discovering how organizations deal with continuous transformation processes. https://www.linkedin.com/in/marco-riccobene-532705202/

Michela Pratola is an IT Enterprise Architect at the Italian broadcaster RAI for two years. She has completed the executive Master's degree in User Experience Design and Information Architecture and is certified in TOGAF. Within the Rai IT department, she mainly focuses on ontology and taxonomy, working on the interoperability of company systems. https://www.linkedin.com/in/michela-pratola/

17:30 – 18:00   **Our journey to a standardized schema in a metadata platform**

What are the key steps and challenges in establishing a standardised schema in metadata platforms?

Sofia Orlova (SRG SSR) Originally from Moscow, I relocated to Zurich at the age of 10 and later pursued a degree in computer science at the University of Zurich. My career began in software engineering, where I gained extensive experience in consulting and freelancing. This diverse background led me to SRF and SRG SSR, where I initially joined as a software developer. My journey within the organization evolved when I took on the role of Product Owner for the metadata platform PDP. Over the last two years, my leadership has been instrumental in steering the platform through significant growth and transformation.

Curdin Capol (SRF) Data Architect at the «Publication Data Platform» of the SRG SSR.
https://www.linkedin.com/in/curdin-capol/

End of Day 2

THURSDAY 14 MARCH 2024

SESSION 11: EXPLORING LLM STRATEGIES – TAILORED APPROACHES AND THE OPEN-SOURCE MOVEMENT
Moderated by Maurizio Montagnuolo (RAI)

09:00 – 09:30   **Adapt LLM to specific tasks: A convenient alternative to the "GPT-for-all" approach?**

On specific tasks, "small" LLMs, if properly fine-tuned, can outperform huge LLMs designed to respond to every user request.
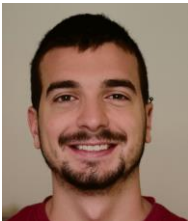
Stefano Scotta (RAI) Having completed my PhD in Mathematics at the Instituto Superior Técnico in Lisbon, I am now a researcher at RAI's R&D Department, specializing in generative AI. My primary focus in this moment is on applying open-source LLMs to address issues in the media industry.

09:30 – 10:00   **Claire LLM model: OpenLLM-France first step to open AI**

The overwhelming success of OpenAI's ChatGPT, whose first version was released one year ago, has led to an undeniable surge of excitement about large language models (LLMs) among researchers and the general public alike. OpenAI's anything-but-open approach to sharing its models or information about training them, however, has led to an equally passionate reaction among those who feel that AI development should be widely accessible and that data usage should be transparent in order to protect the rights of those who have contributed the data and that data – a resource crucial to the development and understanding of AI models – should be shared with the broader research community. A central part of OpenLLM France's mission is to contribute to this momentum by building language models and remaining fully transparent about every step of model training, including the data used for training. Another objective, which we find equally important, is to increase the availability of language models and training data geared to languages other than English and to non-anglophone cultures. Indeed, the majority of the high-profile LLMs available today are trained primarily on English documents coming from anglophone cultures. Only 0.16% of the data used to train LLaMa2 comes from French, for example. After clarifying the notion of an "open source" AI model, the talk will presents our first step to addressing these objectives: the release of Claire French Dialogue Dataset and Model. Claire model and corpus are designed for NLP tasks requiring understanding and generation of spontaneous, oral French dialogue. Claire model and data are available on Hugging Face, code is available on Github. Our current work focus on training "from scratch" a new foundation model, LUCIE 7B, based on a very large corpus of French textual data. The presentation will conclude with a call for contributions to the OpenLLM-Europe community, which aims to bring together European players in open source generative model sector.

Jean-Pierre Lorré (LINAGORA) As the director of research and development at LINGORA Labs, Jean-Pierre Lorré plays a central role in the definition of LINAGORA's global strategy for product development, a vision which has guided him in building the strong team of researchers and engineering experts that serve as the innovative engine behind LINAGORA's products. Jean-Pierre brings industrial and academic expertise in a wide range of topics, including Artificial Intelligence, Automatic Speech Recognition, Natural Language Processing as well as distributed software architecture and open-source models. This broad skill set enables him to tackle important problems in both fundamental research and industrialization, and to play a leading role in over twenty European and French collaborative projects to date. https://twitter.com/jplorre - https://www.linkedin.com/in/jplorre/

Michel-Marie Maudet (LINAGORA)

10:00 – 10:30   **LLM for Media – the new paradigm**

The presentation explores how LLMs (Large Language Models) can transform content analysis, providing additional insights into audience engagement and inform strategic decision-making. Discover how LLMs can enhance metadata generation, thereby improving content organization and delivering an enriched media experience to our audiences.

Fabian Lang (DW) I bring a background in natural language processing, stemming both from academic studies and practical project experience. My primary interest lies in applying NLP within media contexts.

10:30 – 11:00   *Break and demos*

SESSION 12: OPTIMIZING AUDIENCE ENGAGEMENT – FORECASTING, BRANDING, AND MARKETING EFFECTIVENESS

Moderated by Tatjana Mladenovic (BBC)

11:00 – 11:30   **Building a comprehensive BBC TV audience forecasting system**

What happened when we were asked to predict multi-platform audiences for every BBC TV show, from tomorrow to over a year in the future? This talk covers why we do forecasts, the experience of building a hybrid machine learning system, and lessons we learned from stakeholders about how credibility isn't just about accuracy.

Andy Gabey (BBC) I've been at the BBC since 2021, and have been doing what's now called data science since a PhD in 2007. A big chunk of my career has involved building forecasts for real-world uses, making predictions about things that have no interest in behaving straightforwardly - from the atmosphere to man-made catastrophes to audiences. www.linkedin.com/in/andy-gabey

---

11:30 – 12:00   **Data teams: Crafting digital brands for public service media**

Public broadcasting in Germany has faced great challenges in recent years to master the shift from traditional media to digital news and video distribution on multiple platforms. Data is one of the key factors to successfully navigate this transition. This talk provides insights into which methods and technologies can be used to make online media production as user-centric as possible. It takes a look at new processes that need to be established, but also technical challenges when it comes to providing live user data or complex success metrics. How can a traditional newsroom step by step adapt data-based working methods? How can a digital news platform best serve the needs of its users? These are the central questions that are answered in this lightning-talk.

Lilian Dammann is Director of Analytics & Dataplatforms at PubTech, the digital subsidiary of the public broadcasters Bayerischer Rundfunk and Südwestrundfunk. Together with her team, she is responsible for setting up the analytics infrastructure in the cloud and supports journalists and product development teams in the transition to a data-driven work culture. Previously, she worked as a producer and editor, and knows the day-to-day business of public broadcasters firsthand. Lilian Dammann studied production and media management at the University for Television and Film Munich, as well as Philosophy at the University of Philosophy in Munich, SJ. www.linkedin.com/in/lilian-dammann

---

12:00 – 12:30   **Measuring marketing effectiveness at the BBC**

This talk will focus on marketing BBC iPlayer and Sounds content on BBC News and Sport. I'll review the challenges associated with measuring marketing effectiveness using AB tests and hold out groups. I'll then discuss the benefits of using causal inference modelling to passively and continuously measure the our marketing effectiveness.

Matt Crooks (BBC) is a Principal Data Scientist at the BBC. He first started at the BBC in 2018 where he worked on many projects aimed at understanding the BBC's audience. He did a short stint at Typeform before returning to the BBC last year. Since returning he's been working with the marketing teams to measure and improve the effectiveness of Marketing. https://www.linkedin.com/in/dr-matt-crooks-667b7694/

12:30 – 15:00   *Lunch – Demos - Networking*

END of DTS

DEMOS

1	**Mimir - Cloud-native video production and collaboration**	Kristian Kim Eikeland (Mimir Media Tech AS)
2	**Ratings Artist: an ML-based rating prediction framework to incorporate human expertise**	Jens Costers and Adrien Combas (MediaGeniX NG)
3	**Knowledge Graph & LLM, Profile & Aspect based Personalized Media Recommendations**	Alexander Schulze and Ashesh Goplani (Innotrade)
4	**The EBU AI-Hub for Media: Face recognition, Fake News and Meta-Radio applications.**	Alexandre Rouxel, Pierre Fouché, Marin Piguet (EBU)
5	**Deutsche Welle's Approach to a Unified Benchmarking for Evaluating Language Models**	Hala Attig (DW)