

**3RD INTERNATIONAL CONFERENCE ON
AI ML, DATA SCIENCE AND ROBOTICS**
JUNE 26-27, 2023 | Mercure Roma West, Rome, Italy



Miss. Catherine Azam

Gobubble Ltd., Chester, United Kingdom

Addressing Bias in Social Media Content Moderation

Our presentation will focus on the challenges of addressing bias in social media content moderation and how we tackle it using a "Jury of AI models" approach. We recognize that each AI model has its own inherent biases and limitations. Instead of attempting to build an all-knowing AI and which would involve the impossible task of training a model on a perfectly balanced dataset that reflects all religious/ sexual /political groups, as well as all the nuances that make up a human's emotional response to any particular content, we offer a solution that allows users to select their preferred biases while moderating their content. Thus we are not relying on a single AI model, but rather a decision algorithm that sits above the neural networks. Our technology uses Emotion AI + user feedback to create a more accurate and effective content moderation system, eliminating the need for human intervention or filtering. By sharing our innovative approach and success stories with leading AI researchers, we hope to inspire further discussions and advancements in this critical field.

Keywords: AI, NLP, Bias, Social media, Content Moderation

Biography:

Miss. Catherine Azam is a hands-on Data Scientist and Technical Lead with a passion for all things AI, NLP, and Graph Networks. With a background in Applied Mathematics, Statistics, Python, Go, and Linux, she has gained extensive industry experience in the field while working for IBM, SKY, Siemens, various AI startups and a boutique hedge fund. She has built and led engineering teams and is currently scaling her AI content moderation system to moderate billions of real-time messages each day.