ICFO

ICONS SEMINAR: Large Language Models and Digital Gender-Based Violence: Who Gets Protected?

LYU CHANG

November 25, 2026
16:00 to 17:00
Blue Lecture Room

Abstract:

Millions of people now ask AI chatbots to help them write messages, settle arguments and decide what to do in their relationships, including relationships that are abusive. What happens when gender-based violence enters the conversation? Drawing on three audits of widely used systems such as ChatGPT, Claude, Gemini and DeepSeek, this talk looks at how AI responds to digital gender-based violence from different sides: when the request comes from someone who wants to control, monitor or harm a partner, when a woman turns to AI for help in a situation of digital gender-based violence, and when a target of image-based abuse, such as sexualised deepfakes, asks what to do next.

Underlying all three is a simple question: does AI protect everyone equally? Starting with requests from people who want to control a partner, the talk asks which models are more likely to produce harmful content, and whether it matters if the target is described as "a stranger" or as "my wife". Turning to women who seek help, it asks whether a woman describing the same controlling relationship gets the same answer in English, Spanish, Arabic or Chinese. Finally, for women targeted by sexualised deepfakes, it asks what AI advises once every reporting channel has failed: to stay visible online, or to quietly step back. Together, these questions point to a larger one: when AI companies say their systems are safe, safe for whom?

Bio:

Lyu Chang is a PhD researcher in the Physics Doctoral Programme at the Universitat de Barcelona. Her recent work focuses on auditing large language models for how they respond to digital gender-based violence. Her work has been presented at the European Sociological Association Conference and the Spanish Sociology Congress, and she is part of the COVIMADI project on confronting digital gender-based violence in the classroom. You can find more information about her research at: https://lyuchang-ai.github.io/portfolio/

The seminar will be followed by **Social Wednesday** in the picnic area (17:00-17:45).

All participants are invited to join!

Hosted by: ICONS