AI for Good
Global Summit

# AI Standards Exchange

*Challenging the status quo of AI security*

**11 July 2025**
Geneva, Switzerland

IEC ISO ITU

# *The Rising Threat of Hyper-Realistic Generative AI*

"Fake content is becoming indistinguishable from reality"

- Current capabilities:
  - – New models (Veo3, DALL·E 3) generate 8s videos with perfect lip-sync
  - – 98% of people can't spot AI faces in recent MIT tests
- Emerging risks:
  - – Political deepfakes swaying elections (e.g. Biden voice clone robocalls)
  - – $2.5B lost to AI-powered financial scams in 2023

# *AI as the Detective – Current Detection Methods*

"Fighting fire with fire: AI detecting AI"

- Technical approaches:
  - – Spatial forensics: Pixel-level artifact detection (Microsoft Authenticator)
  - – Temporal analysis: Unnatural blinking/gesture patterns (Meta's system)
  - – Multimodal verification: Audio-visual sync checks (Intel FakeCatcher)
- Limitations:
  - – Requires constant model retraining (new forgery techniques emerge weekly)
  - – Fails with high-quality synthetic content (e.g. Sora-generated videos)

# *The inevitable Future – Perfect Digital Forgery*

”When seeing is no longer believing”

- Projected timeline:
  - 2025: >50% of social media videos potentially synthetic (Gartner)
  - 2027: AI passes "Turing Test" for video authenticity
- Fundamental challenge:
  - Both generation and detection use same neural architectures
  - Eventually reaches equilibrium where fakes are perfect copies

# *A Systemic Defense Framework*

”No silver bullet – layered defense required”

- Technical layer:
  - Mandatory watermarking (e.g. C2PA standard)
  - Real-time detection APIs for platforms
- Regulatory layer:
  - China-style deepfake labeling laws
  - Platform liability for unchecked synthetic content
- Social layer:
  - Media literacy programs (teaching "digital skepticism")
  - Verified content channels (like BBC's verified reporter system)