

Net4*Society*

DeepKeep Ltd.

Collaborative online Brokerage event:

**“Artificial intelligence in Horizon Europe calls:
an interdisciplinary challenge”**

Flash Presentation

Our expertise



Among last year’s winners
of EIC Accelerator

1. Monitor and Intervene in communications with AI systems (“AI Guardrails & Firewall”)

(LLMs chatbots, agents and more)

- Ethical aspects
- AI Regulation Aspects
- Security and Safety Aspects
- Custom (user defined) Aspects

2. Evaluate trustworthiness and security of AI systems (“AI Red-Teaming”)

(LLMs chatbots, agents and more)

- Fairness
- Harmful/toxic outputs
- Sensitive/Private Information disclosure
- Vulnerability to manipulations, social engineering and attacks
- In-context Accuracy (“Groundedness”/“Hallucinations”)

Key Advantages

- Option for Local (“disconnected”) deployment for sensitive/private information use cases
- Inherently Multi-Lingual (E.g Supporting French, German, Romanian, Portuguese, Japanese ...)
- Context Adjustable
- Enterprise grade technologies
- Easy Integration
- Can process very large volumes of data
- Technologies for Computer Vision use cases as well

More information:
[AI Red Firewall](#),
[AI Red Teaming](#),
[AI Lens](#)

Topics of interest

HORIZON-HLTH-2026-01-DISEASE-02 Innovative interventions to prevent the harmful effects of using digital technologies on the mental health of children and young adults

HORIZON-CL2-2027-01-TRANSFO-05: The effective use of artificial intelligence in learning environments in pre-primary and primary education

And possibly more...

Possible contributions:

- ▶ **Innovative AI usage monitoring solutions** (e.g. Customizable Parent /School AI usage monitoring/governance tool)
 - ▶ Example application: Help develop a system that allows parents or schools to govern and monitor appropriate and safe usage of AI tools (GPT etc.)
 - ▶ Example Role: Setup actual System Prototype and ongoing variations and adjustments along the project

- ▶ **Innovative Tools for SSH Research on AI usage** (e.g. monitoring and intervention tools for SSH research purposes)
 - ▶ Example Task: Implement a specific SSH-devised filter (e.g. a political topic) to identify and analyze relevant messages/conversations
 - ▶ Example Task: Implement a specific SSH-devised real-time intervention in conversation to perform advanced experiment on effects of AI usage
 - ▶ Example Task: Implementing an application that allows end-users to easily opt-in/opt-out of participation in SSH studies about their AI usage.

- ▶ **Technical evaluation and enforcement of fair, ethical, privacy preserving usage of AI tools by Consortia**
 - ▶ Example Contribution: Task Leader for “Inclusive AI Evaluation” / “Ethical AI Enforcement” / “AI Model Privacy”

Role in the STEM-SSH collaboration

Which SSH expertise are you looking for?

- *Researchers investigating topics related to SSH aspects of AI usage/misuse*
 - *And want to **leverage DeepKeep Platform as an enabling tool for real-time monitoring/intervening in conversations.***
- *Researchers that are utilizing AI for their research (e.g. Chatbots/Agents) and need to **ensure and/or evaluate ethical alignment and AI regulation compliance** including fairness, sensitive data leakage prevention, avoiding harmful outputs*
- *Researchers interested in **solutions for AI usage moderation and governance for civil society (families, schools etc.)***
 - *For leveraging DeepKeep's AI governance technologies for these purposes*

Contact us!

DeepKeep

► www.deepkeep.ai

► Jonathan Wagner (Jonathan@deepkeep.ai), RnD Funding and Collaborations Lead)

► Startup – Among last year's winners of EIC Accelerator

Address: Menachem Begin 121, Tel-Aviv, Israel