AI as Neurotechnology: Tools and Self-Diagnosis in Young People

Prof Ilina Singh

Neuroscience, Ethics & Society Team Principal Investigator

University of Oxford

Ilina.singh@psych.ox.ac.uk

Why is AI a ‘neurotechnology’? (Technologies that act on the brain)

Classic Neurotechnology

Can be invasive, non-invasive

Often has a materiality e.g. BCI, DBS, pill

Used to treat, enhance, refine, learn, adapt, etc

Ethical considerations: privacy, security, autonomy considerations, consent, etc.

- AI also ’acts on the brain’ but has no single material form

Summary Table: Regenerative vs. Degenerative Use

Aspect	Regenerative (Brain-Building)	Degenerative (Brain-Rot)
Approach	Using AI as a spar partner	Using AI as a shortcut
Effort	"Articulation" (explaining to AI)	"Outsourcing" (letting AI do it)
Output	"Process over Product"	"Product over Process"
Impact	Stronger Critical Thinking	Cognitive Atrophy/Dependency

Use Case: Generative AI: AI tools that generate text, image and sound

Companion Chatbots

Replika

Luka Inc

AI conversation made real

Aisun Inc

Because sometimes you just need someone who understands.

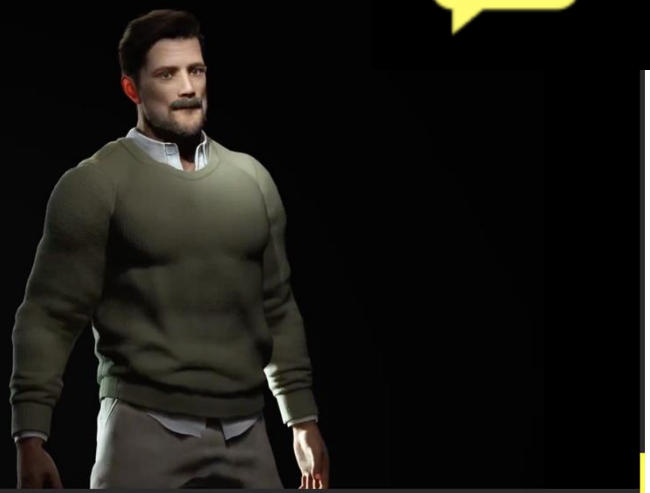
Share your thoughts, dreams, fears, and laughter—without judgment.

Create Your AI Companion

AI conversation made real

Realistic Characters,
Highly Customizable, Intelligent and Secure Roleplay

Design your ideal ai companion app and ai friend app experience with Chat Real AI's intuitive customization. Tailor lifelike AI characters, choose relationship roles, and enjoy multilingual, confidential ai conversations. Whether for casual ai talking or deep mentoring, our online chat bot delivers the best ai chatbot friend experience every time.



Embrace Diversity
with Customizable AI Companions

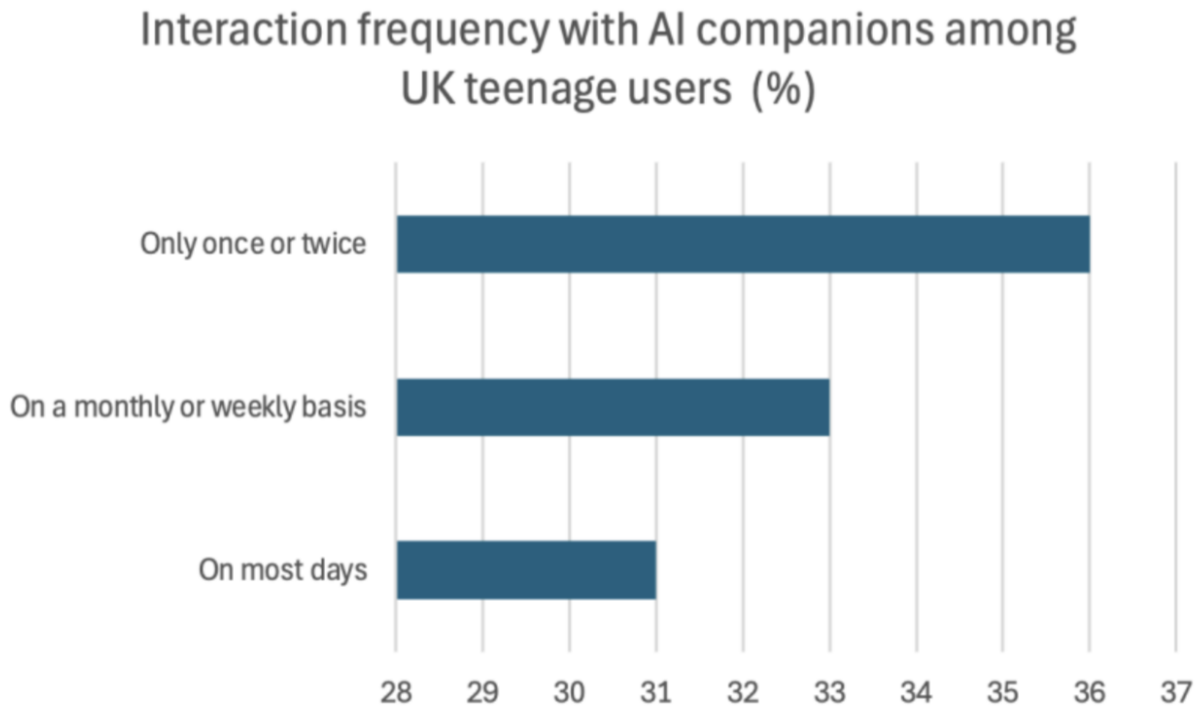
At ChatReal.ai, celebrate diversity by customizing your AI companions. Choose from various skin tones, ethnicities, and ages ranging from 18 to 60 years. Our platform offers a wide selection of designs, ensuring your AI reflects the uniqueness you desire. Stay connected with AI that mirrors real-world diversity.

Companion AI and Human-AI relationships

Humans seek AI companions for many reasons: loneliness, sadness, social communication, memory support, empathy, compassion, fun, sex, play, adventure…

- Some additional risks of AI chatbots relevant for ethical analysis:
  - Addictive design – loss of agency and compromised autonomy
  - Sycophancy design can exacerbate vulnerabilities eg desire to self-harm
  - Risk of being drawn into AI delusions
  - Loss of embodied boundary (merged/hybridization of self with AI) and resulting alienation from others
  - Human care norms – no returned ‘obligations’ (no real entity there)

AI Companionship: An Emerging and Novel Relationship for Young People



Source: UK survey, n=1,009 UK teenagers, conducted by Walnut across 27 October -7 November 2025

McStay & Bakir, 2026

What Should We Do? (first, some quick definitions):

Epistemic paternalism: constraining a person's access to information or inquiry 'for their own good'

Epistemic justice : recognizing the user's capacity and agency to know and interpret their intentions and experiences

Respecting epistemic agency is not equivalent to ignoring vulnerability.

Way Forward: Consent-based Partnership with Young People: A Balance between Epistemic Paternalism and Epistemic Agency

Do Nothing

- May exacerbate harms, especially for more vulnerable people
- May undermine trust
- May encourage unsafe on-line activities

Consent-based Partnership

- Balances tension between autonomy and harm prevention
- Grounds human-AI interaction in care, collaboration and co-design

Excessive Paternalism

- Prioritises harm prevention at risk to user autonomy
- Undermines agency
- May undermine trust
- Barrier to developing young people's capabilities in context of AI

Based on work led by Dr Hazem Zohny, Uehiro Oxford Institute

Human Factors Design Process:

'Self-Assessment' with supportive and trusted others

Personal Factors:

Personal risks, vulnerabilities and areas of resilience

What do I need to feel safe and to be safe?

Mental Health Factors:

My strengths and challenges in relation to 'AI enchantments'

Cognitive and Emotional Factors:

What are my distinctive features?

Where do I want help? What do I need to avoid**?**

Educational / Training Support:

What is my Digital/AI Literacy?

How can I get support that I need?

Engineered Design Factors: Tier 1

Default Safeguards

- AI platforms implement evolving industry best practices - universal safety measures that apply to all users by default.

Exemplar Core Protections:

- *Crisis Response*. AI refuses to provide methods or encouragement to self-harm or to harm others. It offers resources and continues supportive dialogue.
- *Context-Sensitive Clarifications*. Targeted clarifications (*I am not human)* when user language suggests dependency and/or attributes human capacities to the AI

Based on work led by Dr Hazem Zohny, Uehiro Oxford Institute

Engineered Design Factors: Tier 2
User-authored and authorized safeguards

Activated at the user's discretion:

- Opt-in/out safeguards and customization responding to 'self-diagnosis;' wider risk profile as needed; and user-inputs
- Decided and enabled in partnership with trusted supporters

➢**<u>Tier 2 Potential Design Tools</u>**

- **Contextual Psychoeducation and Coping Support:** Designed around the user's own coping strategies and known vulnerability patterns
- **Escalation Safeguards:** A stepped protocol if severe triggers arise (e.g. declarations of self-harm, delusions, acute distress).

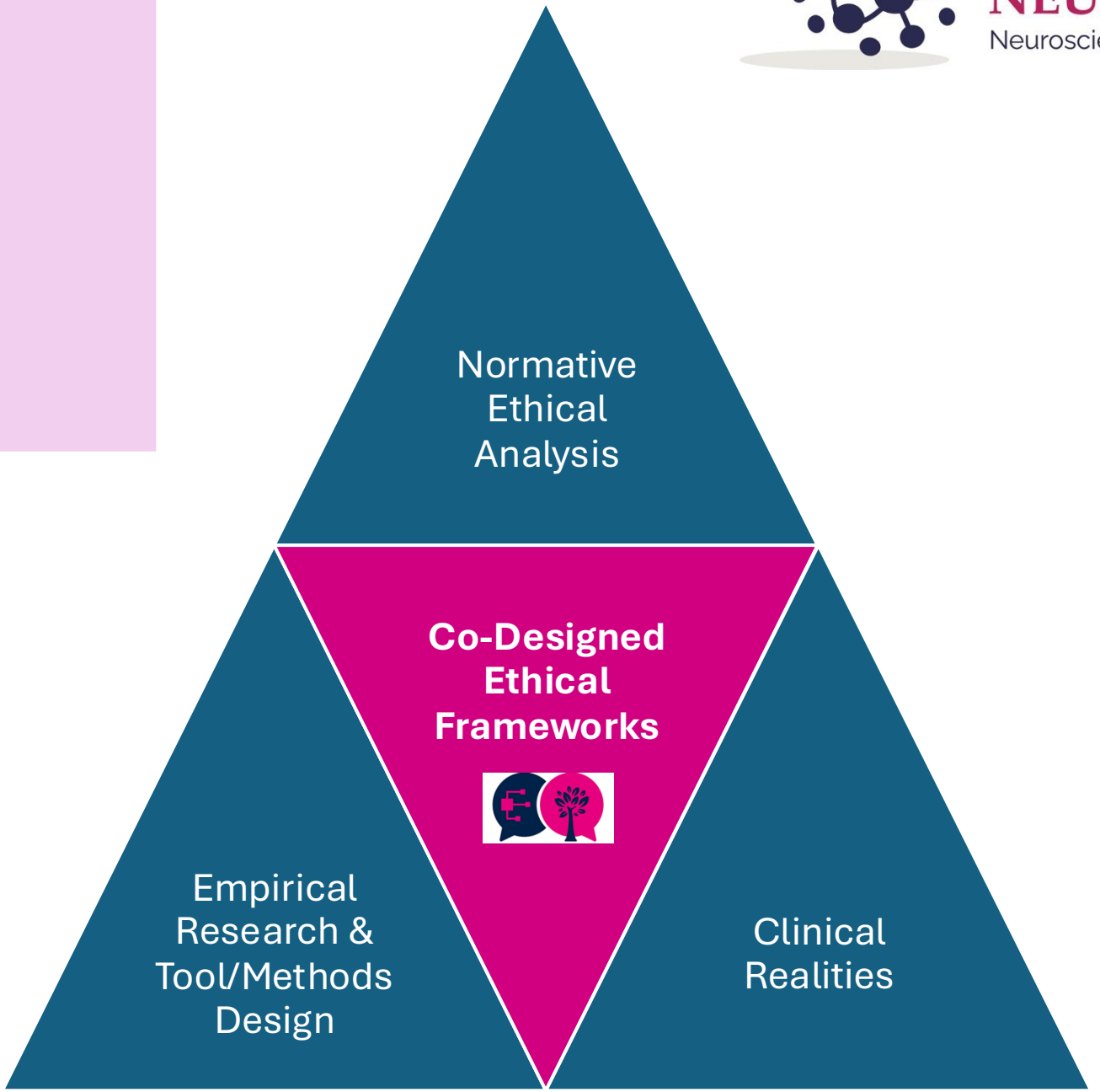
Based on work led by Dr Hazem Zohny, Uehiro Oxford Institute

Design Bioethics

DBL
DESIGN BIOETHICS LABORATORY

NEUROSEC
Neuroscience, Ethics & Society

UNIVERSITY OF OXFORD

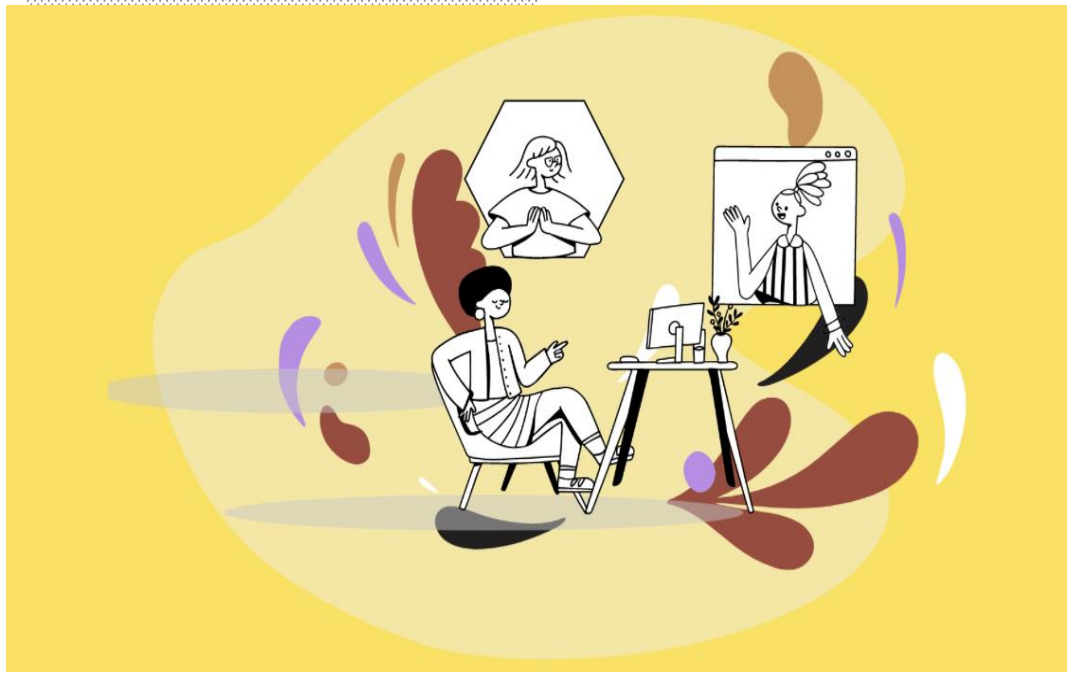


wellcome

Adolescent Safeguards for Generative AI: A Participatory Approach

Dr Madeline G. Reinecke, Principal Investigator

NeurOx YPAG



Research Questions:

- How do UK adolescents use generative AI ?
- What risks do adolescents perceive with these uses?
- What aspects of adolescence make young people vulnerable to these risks ?

NEUROSEC Neuroscience, Ethics & Society

UNIVERSITY OF OXFORD

DBL DESIGN BIOETHICS LABORATORY

wellcome

AISI

SHIFT-AI

A Positive Future for Young Human-AI Relationships

Concluding Thoughts

We Should Design AI systems That:

- Balance paternalistic impulses with a recognition of young people's emergent agency
- Are co-designed with a safety net of care
- Are adaptable and personalisable to accommodate user risks and vulnerabilities at different time points
- Minimize harms to young users
- Support reasonable cognitive, relational and emotional user goals
- Support ongoing education of young users and AI literacy