Lize Alberts (she/her)

Personal Website   lize.alberts@cs.ox.ac.uk

Fourth-year DPhil Candidate in Computer Science at the **University of Oxford**. Former student researcher at **Google UK** working on large language model (LLM) evaluation. **Research Fellow** at Stellenbosch University's Unit for the Ethics of Technology. Unique multi-disciplinary expertise across **STEM** (*Computer Science*), the **Humanities** (*Philosophy, Linguistics, Applied Ethics*), the **Social Sciences** (*Social Anthropology*), and their **interdisciplinary research areas** (*Cognitive Science, Human-Computer Interaction, Responsible Innovation, Computational Linguistics*).

Publications

- Alberts, L., Lyngs, U., & Van Kleek, M. 2024 "Computers as Bad Social Actors: Dark Patterns and Anti-Patterns in Interfaces that Act Socially", to appear in *Proceedings of the 2024 ACM SIGCHI Conference on Computer-Supported Cooperative Work & Social Computing (CSCW), San José, Costa Rica, November 9-13, 2024.* arXiv.2302.04720. DOI

- Gabriel, I., Manzini, A., Keeling, G., Hendricks, L.A., Reiser, V., Iqbal, H., Tomasev, N., Ktena, I., Kenton, Z., Rodriguez, M., El-Sayed, S., Brown, S., Akbulut, C., Trask, A., Hughes, E., Bergman, A. S., Shelby, R., Marchal, N., Griffin, C. ... Manyika, J. 2024. "The Ethics of Advanced AI Assistants". *Google DeepMind.* PDF

- Alberts, L., Keeling, G., & McCroskery, A. 2024. "What Makes for a 'Good' Social Actor? Using Respect as a Lens to Evaluate Interactions with Language Agents", under review at *Minds & Machines.* arXiv.2401.09082. DOI

- Lyngs, U., Lukoff, K., Slovak, P., Inzlicht, M., Freed, M., Andrews, H., Tinsman, C., Csuka., L., Alberts, L., Oldemburgo de Mello, V., Makransky, G., Hornbæk, K., Van Kleek, M., & Shadbolt., N. 2024. "'I finally felt I had the tools to control these urges': Empowering Students to Achieve Their Device Use Goals With the Reduce Digital Distraction Workshop", to appear in *Proceedings of The ACM CHI Conference on Human Factors in Computing Systems, Honolulu, Hawai'i, USA, 11-16 May, 2024.* PDF

- Alberts, L., Lyngs, U., & Lukoff, K. 2024. "Designing for Sustained Motivation: A Review of Self-Determination Theory in Behaviour Change Technologies", accepted for review for a special edition of *Interacting with Computers: Self-Determination Theory in HCI.* arXiv.2402.00121. DOI

- Lukoff, K., Lyngs, U., & Alberts, L. 2022. "Designing to Support Autonomy and Reduce Psychological Reactance in Digital Self-Control Tools", in *Self-Determination Theory in HCI: Shaping a Research Agenda. Workshop at the ACM CHI Conference on Human Factors in Computing Systems (CHI'22).* PDF

- Alberts, L. 2020. "Not Cheating on the Turing Test: Towards Grounded Language Learning in Artificial Intelligence", master's thesis. arXiv.2206.14672. DOI

- Alberts, L. 2019. "Meeting Them Halfway: Altering Language Conventions to Facilitate Human-Robot Interaction". *Stellenbosch Papers in Linguistics Plus*, 56:97-122. DOI

Work Experience

Student Researcher   July-October 2023
Google London, UK
I worked for three months as a student researcher in the Cerebra team at Google. I led a project critically evaluating current alignment approaches and frameworks for Large Language Models (LLMs), taxonomised unexplored risks, and proposed a novel ethical framework centred on duties of respect.

Leader of Royal Institution Masterclass in Computer Science 2023-
The Royal Institution, UK
I have designed a Royal Institution Masterclass on *Dark Patterns* in user interface design, which I have been leading annually for high school children.

Graduate Research Assistant 2020-2021
University of Oxford, UK
I worked as an RA for the EPSRC-funded RoboTIPS project taking a Responsible Innovation approach to the research and development of social robots.

Teaching Assistant and Seminar Leader
- *University of Oxford, UK* 2020-2021
  Software Engineering Dept.: *Interaction Design, Requirements Engineering*
  Computer Science Dept.: *Ethical Computing in Practice, Ethics and Responsible Innovation*
- *Stellenbosch University, ZA* 2018-2019
  I helped administer two Philosophy courses and gave seminars thrice a week for groups of 5 to 80 first/third-years.
- *North-West University, ZA* 2015-2017
  I was TA for several Philosophy and Art History courses and led weekly seminars.

Education

D.Phil. in Computer Science Finalising—Passed Confirmation
University of Oxford, UK
My doctorate employs HCI methods, as well as critical engagement with interdisciplinary literature, to improve how AI/automated systems treat people in interactions. I particularly focus on LLM agents and other systems that behave as social actors. The aims include better supporting user autonomy and wellbeing, as well as anticipating and mitigating risks regarding emerging language technologies.

M.A. (Thesis) in Philosophy Dec 2020
Stellenbosch University, ZA
Type: Full research thesis (140 pages)
Domains: Philosophy of Mind and Language, Cognitive Science, Cognitive and Computational Linguistics, NLP and NLU (theory)
Title: 'Not cheating on the Turing Test: Towards grounded language learning in AI'
Final grade: **Distinction 84%**

B.A. (Hons) in Philosophy Mar 2019
Stellenbosch University, ZA
University of Bristol, UK (4 months Study Abroad)
Bristol: Philosophical Issues of the Physical Sciences, Probability and Rationality
Stellenbosch: Nietzsche, Philosophy of Language Average: **80%**
Dissertation: 'Meeting them halfway: Altering language conventions to facilitate human-robot interaction' (18566 words) Mark: **88%**
Final grade: **Cum Laude 82%**

B.A. in Humanities Mar 2018
North-West University (Potchefstroom), ZA
Majors: **Philosophy**, **History of Art**, **Social Anthropology**
Minors: English, History, Communication Studies
Extra modules: 9 (152 extra credits)
Final grade: **Distinction 82% (Graduated top of class)**

Eldoraigne High School, *ZA*
Final marks: Information Technology (87%), Maths (83%), Design (97%),
English (85%), Afrikaans (89%), Life Orientation (81%), Dramatic Arts (81%)
Nov 2012   Final grade: **Distinction 87%**

**External
Research Talks**

- **Computers as Bad Social Actors**   8 Mar 2024
  Princeton ETHICOM AI & Ethics Seminar, Princeton University, NJ, USA
- **Dark and Anti-Patterns in Social Interfaces**   13 Sep 2023
  Google-Cambridge University Reading Group on Conversational Agents, Google London, UK
- **Dark and Anti-Patterns in Social Interfaces**   10 Sep 2023
  TL/UXR, Google, London, UK
- **Respect as a Value for Human-Computer Interaction**   21 Mar 2022
  Institute for Ethics in AI gathering with Accenture, University of Oxford, UK
- **Meeting Them Halfway: Altering Language Conventions for Human-Computer Interaction**   28 Oct 2019
  Leverhulme Centre for the Future of Intelligence, Cambridge University, UK

**Research
Workshops**

Google's Techno-Moral Scenarios for Responsible Innovation 17-19 Oct 2023
Google Research/Responsible Innovation, Google Amsterdam, the Netherlands
I was invited to join a select panel of experts from academia and industry to exchange views on using responsible innovation to advance a vision of AI that is empowering, meaningful, and beneficial. We practised using a combination of RI methods (e.g., design fiction, speculative design, and Google's own *techno-moral scenario* approach) to imaginatively and reflexively engage with ethical concerns regarding new technologies.

Generative AI and Its Impact on Human Creativity   1 Feb 2024
Google and the Graziano Lab, Princeton University, NJ, US
I was invited to join a select panel of scholars and Googlers to produce a set of responsible and ethical recommendations regarding Gen(erative) AI. We reflected on the nature and processes of human creativity, and how GenAI may affect, and ideally support it.

Self-Determination Theory in HCI: Shaping a Research Agenda 9 May 2022
The ACM CHI Conference on Human Factors in Computing Systems, Hawaii, US
I participated in a CHI workshop with other HCI researchers to explore how self-determination theory may aid the design of user-centred technologies.

Skills

- **Programming:** Python (deep learning/neural networks); Object Pascal; MATLAB (machine learning); Octave; C; C#; HTML; JavaScript; SQL; PHP
- **GenAI-specific research**: Using responsible innovation methods to evaluate ethical concerns surrounding foundation models; familiarity with RLHF approaches and self-correction strategies for LLMs; experience with prompt engineering, fine-tuning, alignment; machine learning/deep learning theory and practice.
- **Empirical research:** Quantitative and qualitative research methods with human participants; thematic analysis; systematic and scoping reviews
- **Applied research:** Integrating literature from theoretical and practical disciplines; critical and conceptual analysis; applied ethics; evaluating the potential risks/impacts of AI technologies in multidisciplinary teams
- **Academic and Creative Writing:** Writing/reviewing academic papers, poetry

- **System and Graphic Design:** Interaction design, requirements engineering, user-centred design, value-sensitive design, speculative design, graphic design

---

Additional Certification

- **Neural Networks and Deep Learning** by deeplearning.ai on Coursera
  Grade: **100%** (Credential ID: XPWXVSS9FEHH)   Dec 2018
- **Diploma in Web Development** by Shaw Academy   Jul 2016

Organisational Experience

Leader of Interdepartmental Reading Group on LLMs   2023-
The University of Oxford, UK
I lead a biweekly reading group on emerging research on LLMs where Oxford University researchers from Computer Science, Engineering and the Oxford Internet Institute (OII) share their work, discuss new developments in the field, and take GenAI courses.

Co-founder and Organiser of RTI Student Network   2020-
The Responsible Technology Institute (RTI), The University of Oxford, UK
The RTI is an international centre of excellence on responsible technology. I co-founded its international student network connecting researchers interested in RI. I help to organise reading groups, expert panel discussions, and work-in-progress seminars.

President of Oxford University CompSoc   2021-2022
The Oxford University Computer Society (CompSoc)
I was the first female president of CompSoc (est. 1978), one of Oxford University's largest and oldest societies for computer science-related interests. I delegated work, engaged with industry sponsors, and helped organise a university-wide hackathon.

Awards and Achievements

Graduate Lighthouse Scholarship   2020
The Responsible Technology Institute and Department of Computer Science, University of Oxford, UK
My doctorate is funded by a highly competitive scholarship awarded by the Responsible Technology Institute.

Awards for Academic Achievement
- *Department of Philosophy, North-West University, ZA*   Mar 2018
  Merit certificate for excellent academic achievement in Philosophy (top of class)
- *School of Social and Gov. Studies, North-West University, ZA*   Mar 2016
  Merit certificate for best academic achievement in Medical Anthropology
- *School of Social and Gov. Studies, North-West University, ZA*   Mar 2016
  Merit certificate for best academic achievement in Urban Anthropology

Member of Golden Key International Honour Society   2016
North-West University (Potchefstroom), ZA
I was invited for being ranked in the top 15% of North-West University's sophomores.