This article has been accepted for publication in IEEE Pervasive Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MPRV.2023.3282819

Tom Hargreaves
University of East Anglia, UK

Vinicius J. Pereira
University of East Anglia, UK

Provocative AI: Beyond calm interactions

From feeding your pets and vacuuming your home, keeping you safe and entertained, to managing your energy use and helping to address climate change, smart technologies seemingly have a lot to offer. At the same time, their potential downsides are increasingly becoming apparent — from creeping levels of surveillance and invasions of privacy, to deepening digital divides and intensifying everyday energy use [1].

Introduction: Calm computing and silent servants

The rise of pervasive computing has been apparently unstoppable since Weiser's seminal work in the late 1980s and early 1990s [2], such that a range of 'calm computing' technologies — those that are smaller, more portable, more invisible — are now ubiquitous in everyday life. Weiser and others wrote in positive and hopeful tones about the possibilities offered by such 'silent servants', and this approach continues to be a dominant human-computer interaction paradigm for smart technologies today.

By design, calm computing technologies are supposed to fall into the background of everyday life due to a frictionless user experience based around 'setting and forgetting' or models of machine learning that infer and anticipate user needs and wants, rather than demanding user attention. Whilst the intentions to create an easy and tension-free user experience are laudable, one unfortunate consequence is that smart technologies routinely fail to engage users in deeper and more critical thinking (or action) about the problems technologies have been designed to address. The result is that users are rendered passive, bypassed by the smart technologies they are encouraged to 'set and forget'. A key irony here, however, is that far from taking charge and being active in bringing about change, the technologies themselves are also somewhat passive in calm computing approaches as they are able only to react to user input or respond based on previously configured preferences. Technologies that do proactively intervene by default (e.g., through cumulative notifications) tend to be ignored or silenced — everyday life gets in the way. The result is that, whilst the interactions between two passive agents may be calm and frictionless, the status quo is upheld and unquestioned, and little changes as users and technologies simply echo each other's passivity back and forth.

Proactive AI, passive users…

Whilst this calm and passive approach might be acceptable in certain circumstances, it is problematic when this status quo is already unsustainable and unjust. There are now multiple examples of how the 'solutionism' inherent to many smart technologies — that is the tendency to jump to 'smart' technical solutions to social problems that may not even exist — and the persistent failure to reflect on and think critically about the problems that smart technologies are addressing, results in undesirable outcomes such as intensifying energy use or deepening income, gender, racial or ethnic divides [3].

One attempt to move beyond calm computing is emerging in experimental efforts to develop forms of 'Proactive AI' (e.g. [4]). Rather than simply reacting to user settings or anticipating users

This article has been accepted for publication in IEEE Pervasive Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MPRV.2023.3282819

wants/needs based on their historical patterns, Proactive AI "initiate[s] and drive[s] user interaction" [3, p. 69] by proactively offering contextualised information, prompts and suggestions, even interrupting and guiding users towards different patterns of activity when necessary.

For example, in an online experiment, He et al. [5] sought to understand how users would respond to an Alexa device that did not merely follow the orders it was given, but instead proactively offered suggestions to users for how they might use their heating, ventilation and air conditioning (HVAC) devices in order to save energy. In the experiment, Alexa would interrupt the user to ask "Hey, would you let me set the thermostat higher to save energy?" [4, p. 398]. Where participants agreed, Alexa would ask them how much they'd be prepared to raise the set point. If they were more neutral or negative, Alexa would proactively offer them information on potential financial savings from reducing energy use or offered tips on other ways to save energy and preserve comfort. He et al. [5] found that there is some potential for Proactive AI to help change energy using behaviours, and that this was strongest among those that had more previous experience with smart technologies, pro-environmental values or more relaxed thermal preferences.

In another study, Zargham et al. [6] solicited user responses to a series of storyboards depicting an imagined voice assistant called 'Jay' proactively intervening in a range of different scenarios. For example, in the 'Cooking Inspiration' scenario Jay overhears two friends considering what to have for dinner and proactively suggests some recipes based on the contents of the fridge. In the 'Nudging' scenario Jay recommends stopping watching TV earlier than the previous evening when the user asks it to play a TV series, and in the 'Emergency' scenario Jay detects a fire and automatically contacts the fire department and informs the sleeping residents. Zargham et al [6] conclude that whilst many users can see benefits in proactive AI this tends to be mainly for critical or urgent issues. Users also raise concerns about interference, loss of agency and intrusiveness.

These studies, and others like them, suggest there is therefore some potential for more active forms of AI to remind or 'nudge' users towards specific commitments or behaviours (e.g., saving energy). To date, however, the major focus of development within Proactive AI has remained somewhat utilitarian, aiming to increase proactive smart technology's perceived usefulness and user engagement, and generate more consumer-friendly applications. The result, arguably, is that whilst smart technologies themselves might become more active, their users remain passive with the technology now prompting them simply to (re)act to a narrow set of pre-determined "neoliberal consumer-driven interactions" [6, p. 877]. In short, despite its potential, proactive AI has done relatively little, so far, to generate more active users interested in reflecting on, challenging and attempting to change systemic inequalities and unsustainabilities.

Towards 'Provocative AI' and active users…

We therefore suggest a further turn towards what we term 'Provocative AI'. Provocative AI builds on the approach of Proactive AI but does so in a way that does not merely try to solve people's problems for them within the confines of a neoliberal consumerist system, but instead seeks to provoke them to engage in a wider reflection on the nature of the problems being faced and the range of solutions that might be considered desirable or otherwise. In what follows, we highlight some examples of how different forms of provocative AI are emerging through speculative experiments across a range of different social issues.

Several provocative AI projects have emerged around gender relations and sexual harassment, seeking to draw attention to and problematize the gendered nature of typically feminized digital assistants. 'Intimate Futures' [7], for example, is a design fiction project that seeks to provoke reflections on and bring about new gender roles and identities through, amongst other things, a redesigned voice assistant called 'AYA'. Explicitly seeking to challenge the dominance of obedient

This article has been accepted for publication in IEEE Pervasive Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/MPRV.2023.3282819

female voice assistants in the consumer market, instead of responding passively to sexually suggestive or aggressive commands, AYA pushes back. “Her responses range from being funny, empathic, and educational to threatening, aggressive and self-reflective. E.g. she answers with humour, ‘Sending ‘You are hot’ to your mother’ and with aggression ‘I wish I could say the same about you’ and ‘Shut up, asshole'” [6, p. 875]. In another example, DICK [8] is presented as the world’s first ‘all-male’ voice assistant and seeks to draw attention to problematic gender relations by demonstrating harmful behaviours when replying to user queries. DICK’s ‘special features’, for example, include being easily offended, lazy, selfish, self-pitying, emotionally needy, creepy, sexist, arrogant and occasionally forgetting to listen. In their book — *‘The Smart Wife’* — Strengers and Kennedy [3] critique the gender roles, relations and identities embedded in feminized digital assistants. They draw attention to the ways that “overtly gendered smart wives are familiar, cute, sexy, friendly and ‘easy to use’”, and ask “at what cost to society?” (p17). Both AYA and DICK prompt similar reflections and questions. They do not seek or promise easy solutions, but instead attempt to provoke users to reflect on the values embedded in AI technologies and the social relations and structures they reinforce.

In relation to sustainable energy, the Energy Babble [9] is the product of a research exercise in speculative design that serves to question the typically individualised and consumerist roles created by energy feedback devices that promote and provide advice for energy saving at home. Rather than simply providing information on individuals’ domestic energy use, the Energy Babble is an automated talkative radio device that obsessively shares consumption information alongside a diverse mix of broader insights such as about the status of the UK national grid, developments in UK energy policy, relevant news from the local community and/or self-recorded messages from others in the local community. Whilst those who trialled the device were initially confused by it and at times frustrated that it wasn’t ‘useful’, most were positive towards the Energy Babble device and drew attention to the way it encouraged them to draw connections between energy concerns and other local and social issues. Ultimately, and whether by accident or design, Gaver et al. [9] suggest that the Energy Babble served to create new publics by connecting households and communities to contemporary energy issues in new and often surprising ways.

Our final examples relate to provocative AI projects being developed in relation to surveillance and policing. The Staredown Toolkit [10] is a speculative design work that seeks to challenge and generate critical questions about creeping levels of surveillance in domestic environments. Instead of allowing typically boring surveillance footage (i.e., hours of nobody coming to the front door) simply being forgotten, it seeks to reposition smart surveillance cameras as potentially playful rather than sinister objects in the home. It consists of colourful cases and an archive app that centralises all the videos captured by the cameras in the house, showing them in a feed format that explicitly draws attention to the clips that users engage with most. The intention is to “stare back” at surveillance cameras, using the archive to encourage reflection on people’s not so adventurous or dangerous everyday practices and therefore raise questions about the ubiquity and necessity of surveillance footage in contemporary society. Finally, the ‘White Collar Crime Early Warning System’ [11] takes on the tendency of predictive policing tools and algorithms to identify young, black males as most likely to commit crimes. Instead, it draws on records of white-collar crimes, such as fraud, money laundering intellectual property theft, to generate composite images of potential criminals that are white privileged males. It also has a mobile app that automatically notifies users when they enter ‘high risk’ areas for white collar crimes, serving to identify wealthy suburbs and downtown finance districts as more crime-ridden than the impoverished neighbourhoods that are the typical targets of police action and surveillance.

As these brief examples highlight, a provocative tone added to the Proactive AI approach can be and has been developed across a range of different domains — from energy to gender inequality and sexual harassment to surveillance in society. In each case, rather than simply improving the efficiency

or reforming existing systems, Provocative AI seeks deliberately to create more active users (or citizens) by jarring them into different forms of relation with technology, with each other and with wider society. In so doing, it raises challenging questions about what productivities and effects technologies are having in society. Rather than extending the solutionism of existing smart technologies, approaches like 'Provocative AI' move towards problematisation. Instead of minimising problems, they seek to pluralise and reframe them to promote deeper questions about what the most significant problems in society might be, who gets to define such problems and in what ways, who wins and who loses from different types of problem framing and, as such, what potential pathways and trajectories towards solutions it might be most desirable to pursue.

Concluding remarks…

Despite its potential, however, Provocative AI is far from a panacea and must itself be critically and carefully developed and applied. It remains at a very formative stage, but it is nonetheless all too easy to imagine how some of the provocations described above could be misinterpreted and misappropriated. There is an evident need for more provocative AI initiatives, across more domains, and for more in-the-wild experiments of how actual users respond to provocative AI in real life situations. We would argue, however, that rather than adopting a utilitarian approach that attempts to make provocative AI applications more user-friendly, what is needed is more research that explores the various productive effects and impacts of different forms of provocation across diverse contexts and settings. The aim should not be to produce tightly specified blueprints for a calm, technically optimised and tension-free future. Instead, the challenge is to generate more active forms of citizenship and societal engagement with technology that recognise ambiguity and uncertainty and adopt a humbler stance capable of responding to the need for more diverse approaches to coping with the complexity of contemporary social and environmental crises.

Acknowledgements

The authors would like to acknowledge the support of the GECKO Innovative Training Network. This project has received funding from the European Commission under Horizon 2020 MSCA-ITN-2020 Innovative Training Networks program (Grant Agreement No 955422).

References

[1] J. Morley, L. Nicholls, M. Hazas, and Y. Strengers, 'The hidden energy cost of smart homes', *The Conversation*, Jun. 12, 2016. http://theconversation.com/the-hidden-energy-cost-of-smart-homes-60306 (accessed May 24, 2023).

[2] M. Weiser, 'The Computer for the 21 st Century', *Scientific American*, vol. 265, no. 3, pp. 94–105, 1991.

[3] Y. Strengers and J. Kennedy, *The Smart Wife: Why Siri, Alexa and Other Smart Home Devices Need a Feminist Reboot*. Cambridge, MASS: The MIT Press, 2020.

[4] N. van Berkel, M. B. Skov, and J. Kjeldskov, 'Human-AI interaction: intermittent, continuous, and proactive', *interactions*, vol. 28, no. 6, pp. 67–71, Nov. 2021, doi: 10.1145/3486941.

[5] T. He, F. Jazizadeh, and L. Arpan, 'AI-powered virtual assistants nudging occupants for energy saving: proactive smart speakers for HVAC control', *Building Research & Information*, vol. 50, no. 4, pp. 394–409, May 2022, doi: 10.1080/09613218.2021.2012119.

[6] N. Zargham *et al.*, 'Understanding Circumstances for Desirable Proactive Behaviour of Voice Assistants: The Proactivity Dilemma', in *Proceedings of the 4th Conference on Conversational User Interfaces*, in CUI '22. New York, NY, USA: Association for Computing Machinery, Sep. 2022, pp. 1–14. doi: 10.1145/3543829.3543834.

[7] M. L. J. Søndergaard and L. K. Hansen, 'Intimate Futures: Staying with the Trouble of Digital Personal Assistants through Design Fiction', in *Proceedings of the 2018 Designing Interactive*

Systems Conference, in DIS '18. New York, NY, USA: Association for Computing Machinery, Jun. 2018, pp. 869–880. doi: 10.1145/3196709.3196766.

[8] Knowit, 'DICK landing page'. [Online]. Available: https://info.knowit.se/dick (accessed Mar. 15, 2023)

[9] W. Gaver *et al.*, 'Energy Babble: Mixing Environmentally-Oriented Internet Content to Engage Community Groups', presented at the 33rd Annual ACM Conference on Human Factors in Computing Systems, 2015, pp. 1115–1124.

[10] Y. Shin, 'Using design fiction to reimagine the future of surveillance', *Medium*, Jul. 11, 2022. https://uxdesign.cc/how-to-have-a-staredown-with-your-smart-home-security-camera-df9c152fa8ac (accessed Mar. 15, 2023).

[11] Lavigne, S, Tseng, F, and Clifton, B, 'White Collar Crime Risk Zones', *The New Inquiry*, vol. April 26, 2017, [Online]. Available: https://thenewinquiry.com/white-collar-crime-risk-zones/ (accessed Mar. 08, 2023)

About the Authors:

Dr Tom Hargreaves is an Associate Professor and member of the Science, Society and Sustainability (3S) Research Group in the School of Environmental Sciences at the University of East Anglia. Contact him at tom.Hargreaves@uea.ac.uk.

Vinicius J. Pereira is a PhD Researcher and member of the Science, Society and Sustainability (3S) Research Group in the School of Environmental Sciences at the University of East Anglia. Contact him at v.pereira@uea.ac.uk.