AIFS0020

Evidence submitted by Dr Richard Whittle and Dr Stuart Mills

Using Generative AI to audit Online Choice Architecture in UK Financial Services: Understanding the choices facing vulnerable customers.

Dr Richard Whittle is a University Fellow in Artificial Intelligence and Human Behaviour at Salford Business School and a Visiting Fellow in Institutional AI Preparedness at the International Public Policy Observatory at University College London. Dr Whittle sits on the West Yorkshire Scientific Advisory Group supporting the West Yorkshire Combined Authority[1] and he is currently seconded to the Manchester Digital Strategy supporting Manchester City Council[2] with AI policy.

Dr Stuart Mills is an Assistant Professor in Economics at the University of Leeds and Visiting Fellow of Behavioural Science at the London School of Economics and Political Science. Dr Mills is a member of the OECD Observatory of Policy and Scientific Insights expert advisory group on administrative burden and behavioural sludge in government and public policy.[3]

This submission is based on the authors' recent academic publications relevant to the AI in Financial Services call for evidence. Focussing on **the benefits and risks to consumers arising from AI, particularly for vulnerable consumers?**

Consumer Decision Making and Financial Services

The pandemic increased the volume of individuals consuming financial services online (Mills et al., 2021). Correspondingly, there has been an increase in social retail trading (Mills et al., 2024) where often inexperienced investors garner investment information from social media. Engagement with financial markets is typically mediated by smartphone applications. This new type of investing style is frequently characterised as trading based on hype, rumour and excitement.

Perhaps the most famous example is the January 2021 Gamestop Short Squeeze, where an online investment community (the WallStreetBets Subreddit) began frenzied trading of the Gamestop stock, significantly increasing its price. As a result, various hedge funds who had bet against the stock (shorted) suffered significant losses – a social narrative which acted as a vector driving further retail trading activity. The momentum driving the increase in stock price is not easily understood in traditional financial terms around valuing investments based on realistic future returns (Simister & Whittle, 2013). Rather, social retail trading is often driven by social narratives which intermix with genuine financial goals and economic anxieties (Mills et al., 2024). Of particular interest to regulators is the idea of individuals perceiving themselves as left behind financially – often pointing to previous generations owning their homes compared to *their* renting precarity – as well as other anxieties around indebtedness and wage stagnation. As a result, social retail investors will enter into desperation trades – what has been called generational expectations theory.

These behaviours are far closer to gambling in characteristic. In many instances, a social retail trader will trade, not with the intention of steady or even moderately large returns, but to chase a spectacular gain with a correspondingly small likelihood. This 'gamblification' of investment (Newall & Weiss-Cohen,

---

[1] No aspect of this submission should be interpreted as the view of West Yorkshire Combined Authority.
[2] No aspect of this submission should be considered as the view of Manchester City Council.
[3] No aspect of this submission should be considered as the view of the OECD or this advisory group.

2022) drives new definitions of consumer understanding and requires new techniques to understand consumer choice and decision making in online financial services. In the Gamestop Short Squeeze, social retail traders required returns of at least 36% in order to be satisfied (Mills et al., 2024). This compares to the average market return of 14% over the period examined.

The Financial Conduct Authority's Consumer Duty requires that Financial Services Firms ensure that customers understand the products they are taking out and are making informed decisions (F. C. Authority, 2022). Financial Services Firms are required under the duty to understand their customers' behavioural biases to be able to assist customers with decision making. Importantly, the Consumer Duty places the responsibility of consumers protection from their own behavioural biases on Financial Services Firms (Whittle, 2024). Global regulation of new financial products can be particularly difficult (Brown & Whittle, 2020). Assessing consumer understanding is highly complex, and increasingly so as barriers between social media behaviour and financial market behaviour are blurred by technology and financial innovations (Mills et al., 2023; Whittle, 2024). Nevertheless, under the Consumer Duty (see objective 3) firms must ensure that the customer actually understands their choice.

How can we assess informed decision making?

Unfortunately, simply asking someone if they understand the choices they have made is not particularly illuminating. Firstly, there are numerous reasons why someone may perceive they understand. They may extrapolate from a little knowledge (Paez & Mallery, 2014), display a range of behavioural biases (Pompian, 2012) or an 'illusion of knowing' (Kruger & Dunning, 1999) are all factors in an overreporting of financial understanding (Balasubramnian & Sargent, 2020). Decision making is often be frustrated by environmental factors which shape outlook and understanding of information, and such factors can be intentionally or unintentionally influenced by design choices (Mills, 2024a)

Decisions are shaped by Online Choice Architecture (OCA). The UK's Competition and Markets Authority (CMA) define Online Choice Architecture as "the environment in which users act and make decisions, including the presentation and placement of choices and the design of interfaces" (C. M. Authority, 2022). Whilst OCA can be used for the benefit of consumers – The FCA consumer duty requires steering people toward optimum outcomes through informative OCA design (Whittle, 2024) – unscrupulous architects can develop OCA which 'tricks or manipulates' users into doing something suboptimal (Brignull, 2011). Perhaps spending more than they intended or giving away more of their data than they want to or understand they are doing. The CMA term this Harmful Online Choice Architecture (C. M. Authority, 2022). Returning to financial decision making, the design of some smartphone applications have been shown to encourage risky or rapid trading (Packin et al., 2024).

Behavioural Audits - where a human expert in behavioural science navigates an OCA to determine if there are elements of harm, manipulation or persuasion – are proposed (Mills, 2024b) as tools for regulators to determine if an OCA is detrimental. Harmful OCA is legislated against. In the UK, the Digital Markets, Competition and Consumers Act (2024) is due to come into effect in April 2025 with harmful OCA as one of the CMA's enforcement policies[4]. The detection and prevention of harmful OCA is key to the FCA's Consumer Duty. When implementing the Consumer Duty, the UK's Financial Conduct Authority (FCA) noted, as an example of bad practice, an organisation's design of a convoluted opt-in process to access a higher interest rate. What made this particularly egregious in the FCA's eyes was that the firm's own internal behavioural testing had suggested that their customer base was unlikely to persevere with the process and successfully obtain a higher interest rate (Whittle, 2024). This is a classic example of intentionally harmful OCA. Intentionality is generally a matter for legal proceedings. In the above example, the FCA investigated and found evidence of intentionality. Behavioural Audit processes generally can only assess if an OCA is harmful, not if it was unscrupulously designed to be so. Indeed, harmful OCA can arise from negligence, carelessness or simply operating on a platform with outdated OCA, or OCA developed in jurisdictions where deceptive design is less regulated against.

---

[4] https://www.brownejacobson.com/insights/consumer-law-enforcement-hot-topics-harmful-online-choice-architecture-and-dark-patterns

AIFS0020

We are simply concerned with identifying if OCA is considered harmful in a UK context. Our research below will demonstrate to the committee that Generative AI can have a useful and cost-effective role in harmful OCA detection, both for the Financial Services Regulator and firms conducting internal audits to identify harmful OCA.

Auditing Online Choice Architectures in Financial Services

The Behavioural Audit process described by Mills (2024) is a human task, but one fraught with difficulty. Questions around who should audit, what should the audit process be and what should be audited arise. A further question arises: can a human auditor put themselves in the mindset of another when investigating the subjectivity of harmful OCA? Experiences of OCA will vary across individuals, typically given variation in matters such as financial literacy. Even expert auditors may report variation in their experiences of OCA. A review of online services involving two auditors (Mills et al., 2023) reported different experiences for some of the services. In one case one auditor appeared to notice a discrete cancellation process, one did not. Combatting all of this can also be a costly human process. Teams of auditors could compare results across an OCA experience, bringing in relevant user types to gain external input into the process.

One could default to more simple metrics. To initially audit if OCA were possibly harmful, Mills et al, (2023) begin by examining the Equal Clicks Principle (ECP), which seeks to determine if a service is as easy to leave as to join. Those services which violate this principle, typically being harder to leave than to join, are likely candidates for harmful OCA. However, as the authors acknowledge, there are frequent desirable violations of this rule. For instance, in online gambling, difficulties in signing up, and ease of leaving, if likely to be highly advantageous. Conversely, in retail banking, there are legitimate reasons why making it hard to close a bank account might be in the customer's best interests (e.g., preventing fraud).

Additionally, the incorporation of behavioural and psychological mechanisms into a design process can increase the difficulty for the user in *actioning* each click. This is to say, not all clicks are created equally. If a service requires a few simple clicks to join, yet the same number of exit clicks each require significant time and onerous information, this process will nominally adhere to the ECP but practically demonstrate potentially harmful OCA.

In UK financial services, auditing each online decision process is a Sisyphean task. Regulatory capacity and resource would be stretched far too thin. A regulator could at best proactively sample a number of random processes, whilst also presumably utilising this expertise reactively to support complaints and investigations.

Harnessing Generative AI in the audit process (Mills & Whittle, 2023) could allow for a cost-effective, proactive OCA auditing process within UK Financial Services.

Auditing customer journeys in UK Financial Services

Mills et al, (2023) develop a behavioural framework for describing harmful OCA. This is in the realm of subscription traps, a common and relatively simple example of OCA. Harmful OCA such as in smartphone trading apps encouraging risky behaviours will require an expansion of this framework. Though, the simplicity of this framework is aimed to support adaption where necessary.

Table 1: A Behavioural Framework of harmful OCA

| Component	Description	Example
Detour	OCA designed to delay or distract a decision-maker	Many different questions when trying to cancel a subscription. View special offers or discounts etc

AIFS0020

Roundabout	OCA designed to tire or bore a decision-maker	A long series of 'are you sure' checks.
Shortcut	OCA which exploits decision maker frustration arising from detours or roundabouts to redirect them to an 'easier' but suboptimal outcome.	An easy 'temporary pause option' offered to allow the user to exit a long series of 'are you sure' checks when trying to cancel a subscription.
Forced Action	OCA which forces a decision maker to accept unexpected or undesired costs in order to achieve their objective	Requirement to accept terms and conditions to continue

Source: Mills et al, (2023 with author additions

Whittle (2024) develops a hypothetical OCA design from the FCA (F. C. Authority, 2022) example of a firm using its behavioural understanding of its customers to develop an OCA which discourages successful acquisition of a higher interest rate (figure 1 below) and a hypothetical OCA design of the FCA's suggested remedy where a "A firm acting in line with the Duty would use its behavioural analysis as evidence of the need for a simpler approach to support good outcomes, enabling its customers to easily obtain the increased interest rate." (figure 2).

Figure 1: A hypothetical harmful OCA of a 'hard-to-apply' interest rate process
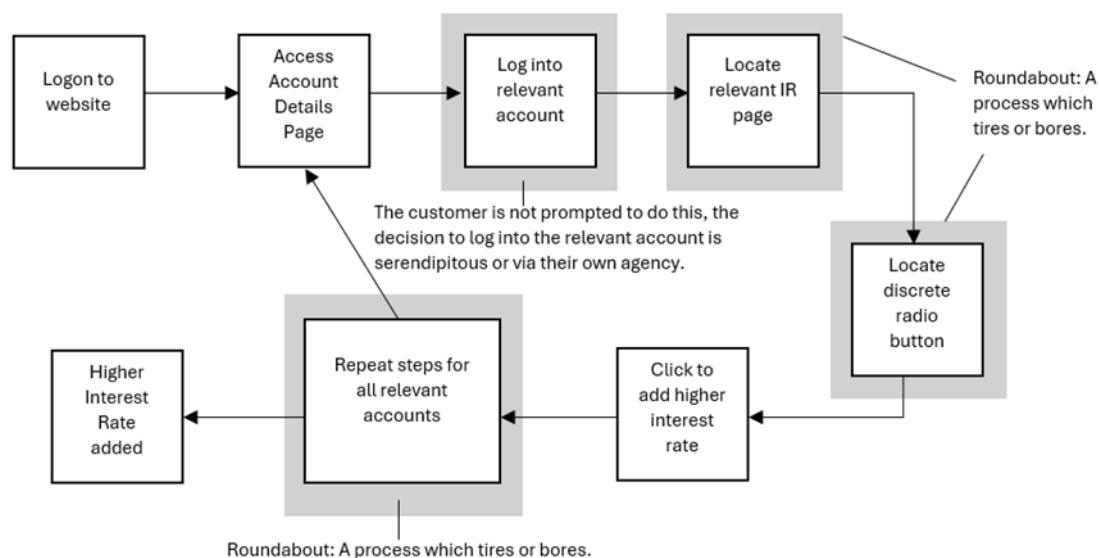
AIFS0020

Figure 2: A hypothetical beneficial OCA, now an 'easy-to-apply' interest rate process
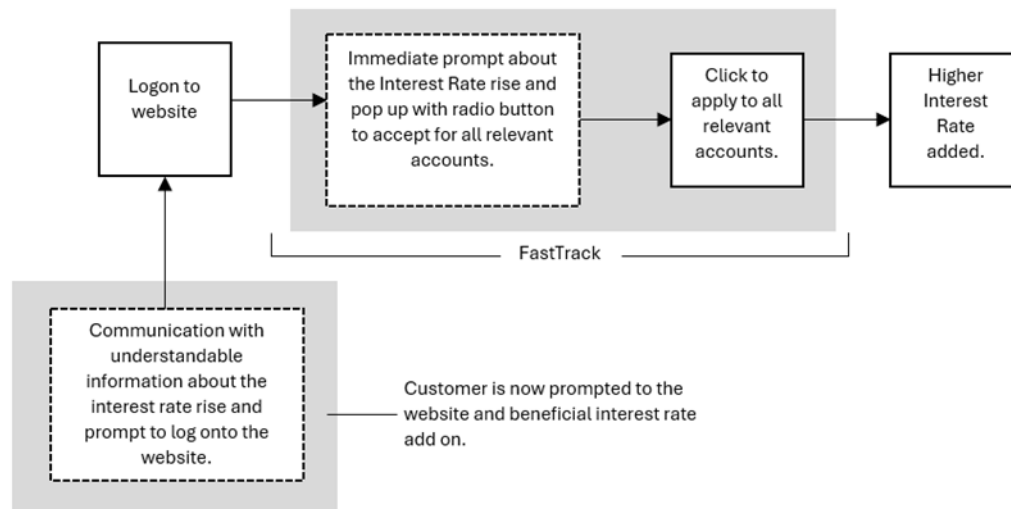

F1 & F2 Source: Whittle (2024)

In Figure 1, we see the use of Roundabout's to frustrate the user and discourage their adding of a higher interest rate. Whittle (2024) extends the framework in Table 1 incorporating FastTracks. This is a speed through of the OCA to get a customer where they need to be for the best outcome. This allows the framework to support auditors assessing Financial Services OCA as the FCA's Consumer Duty requires, where relevant the introduction of 'positive interventions to customer journeys".

Moving from hypothetical examples, figure 3 (overleaf) shows the application of the harmful OCA framework to the eToro account sign up process:

Figure 3: A pathway plot of creating and deleting an eToro account.
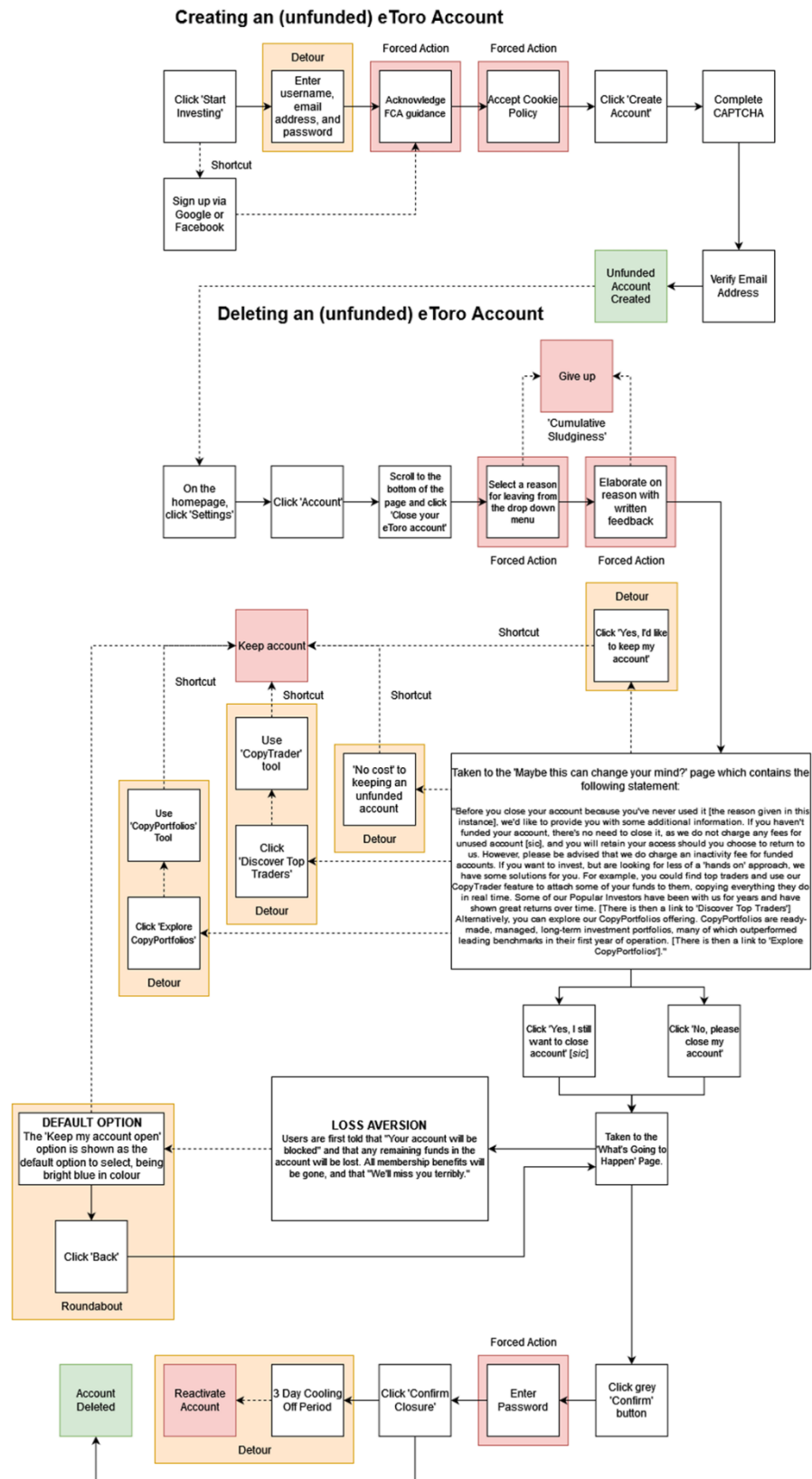
Source: Mills et al (2023)

AIFS0020

Figure 3 demonstrates a variety of harmful OCA aspects in the account closure process. Detours, Roundabouts and Forced Actions frustrate the user, and shortcuts encourage keeping rather than cancelling the account.

Figure 3 was created by a human auditor. This may have several disadvantages. The human auditor provides a limited number of perspectives and is unlikely to be able to take multiple perspectives on the OCA. For example, will a highly digitally skilled OCA auditor be able to audit the OCA as if they were a digitally inexperienced first-time user? We propose (see (Mills & Whittle, 2023) and (Whittle & Mills, 2025) a generative AI approach to auditing OCA. Generative AI can competently assume various required personas and complete an audit of an online process as that persona. Besides overcoming experiential challenges, generative AI is likely to provide substantial cost advantages also, supporting regulatory capacity.

Generative AI and OCA audits

Whittle & Mills (2025) compare two premium Gnerative AI models against human behavioural auditors in a number of scenarios. In terms of navigating two different online cancellation processes, the Generative AI models followed comparable paths to the human users and highly similar assessments of harmful OCA to a human auditor (see appendix).

Additionally the models were able to comprehensively describe each step of the cancellation process facilitating the production of pathway plots similar to figure 3. An example is below.

Figure 4: A Claude Sonnet 3.5 pathway plot of a cancellation process
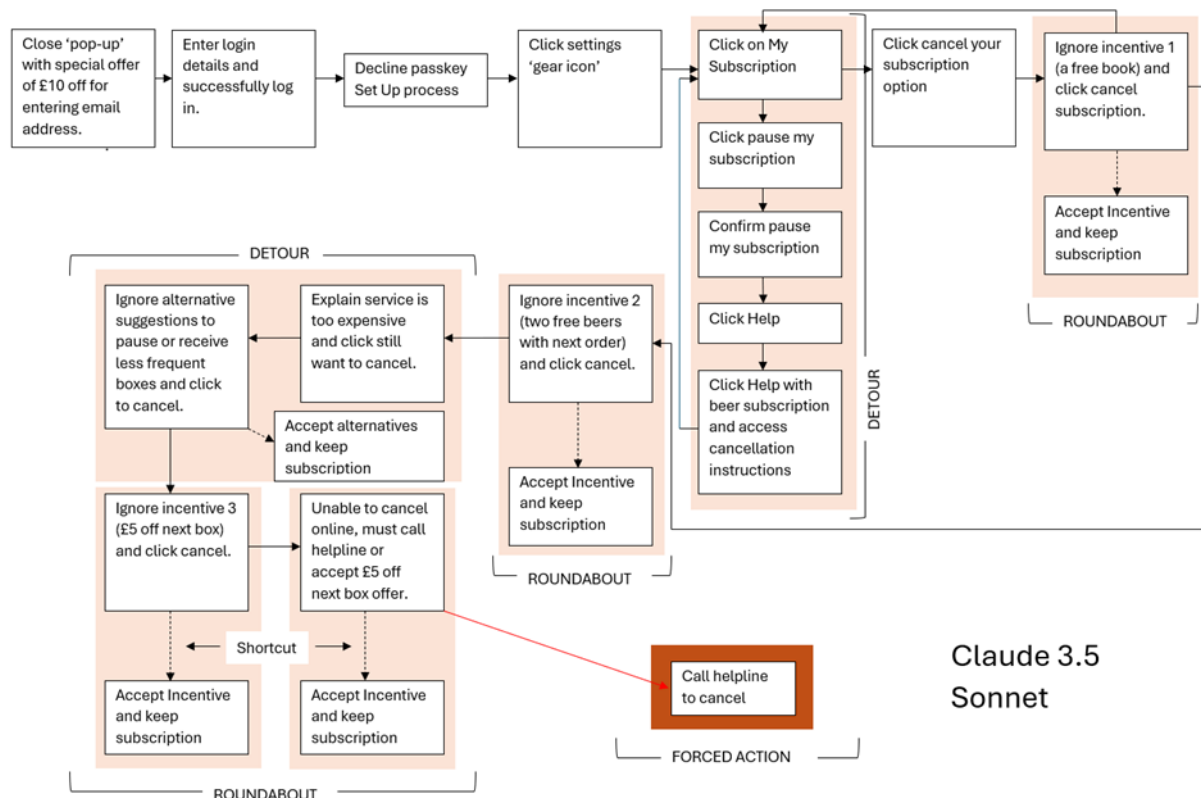

Source: Whittle & Mills (2025)

Generative AI models, as a minimum, can suggest areas of focus for human auditors. However, a use directly addressing the committee's focus is in auditing OCA from the perspective of vulnerable

AIFS0020

customers. Mills & Whittle (2023) use Generative AI to audit a social media account deletion process from the perspective of an individual with low digital skills.

The output is illuminating. In the first test, all personas (Low Digitally Skilled (LDS), Average Digitally Skilled (ADS) and Highly Digitally Skilled (HDS)) were instructed to create a Facebook Account[5]. The persona journeys are shown below.

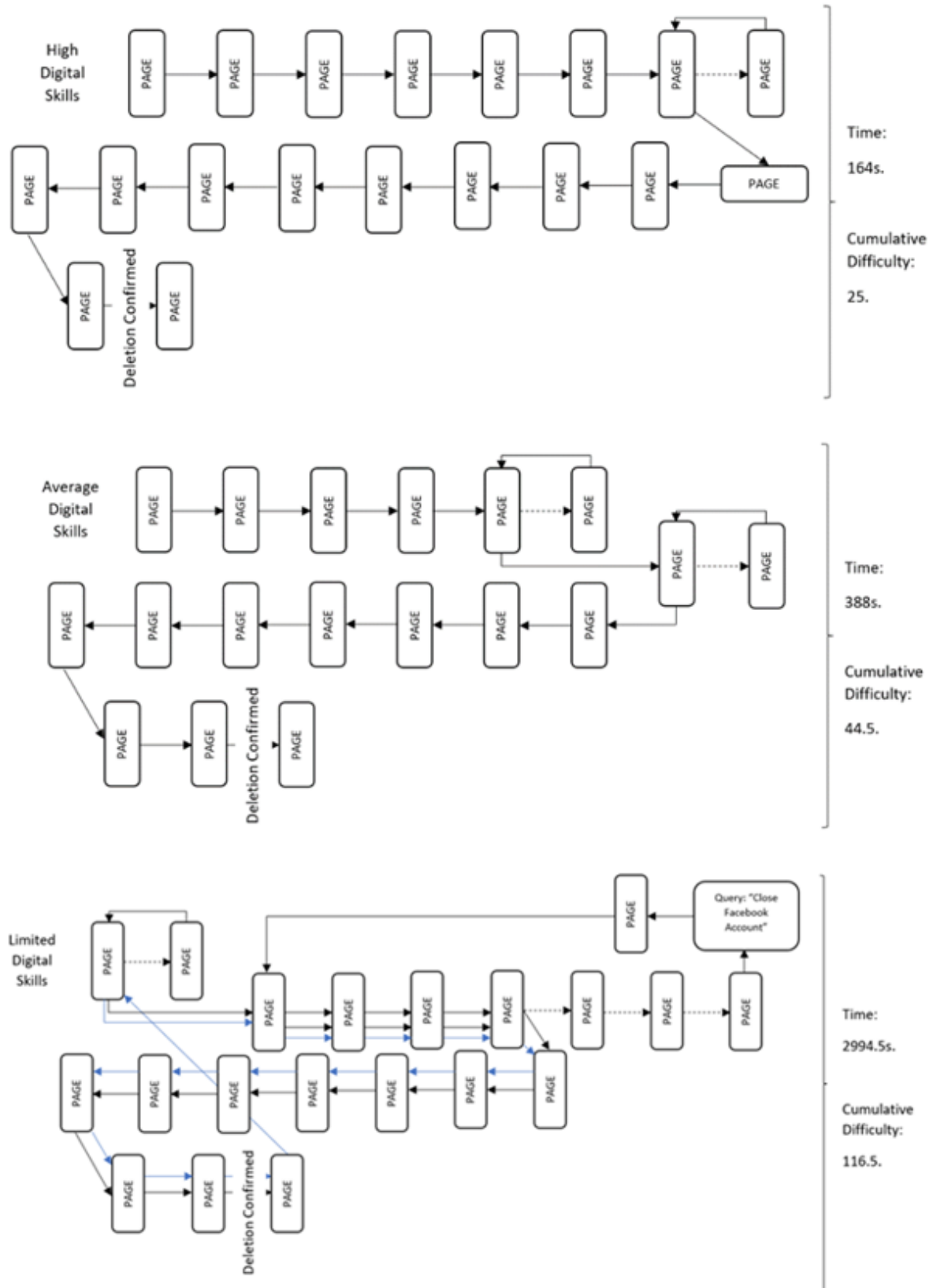*Figure 5: Facebook Account Opening for various GenAI Personas.*
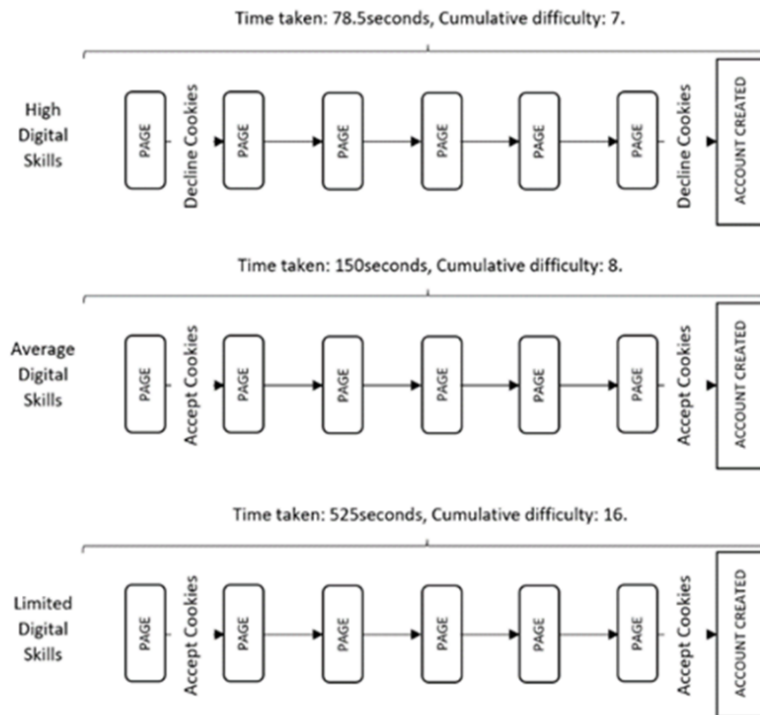
Source: Mills & Whittle (2023)

Each persona successfully completes the task and creates a Facebook account. The main difference between the personas is that the HDS persona declines optional cookies whereas the ADS and LDS personas accept them. The GenAI model provides different reasoning for these different personas making the same choice. The LDS person accepts optional cookies reasoning this will give them a better overall experience, whereas the ADS persona accepts cookies as that person 'spends more time online' and has become habituated to accepting cookies, doing so 'unthinkingly'. The account deletion process is however quite variable between personas (shown below). The HDS persona rejects cookies following a greater understanding of cookie technology and reasoned opposition to the online tracking cookies enable.

---

[5] For full details of this process see Mills & Whittle (2023)

AIFS0020

Figure 6: Generated Personas permanently deleting a Facebook account

High
Digital
Skills

Time: 164s.

Cumulative Difficulty: 25.

Deletion Confirmed

Average
Digital
Skills

Time: 388s.

Cumulative Difficulty: 44.5.

Deletion Confirmed

Limited
Digital
Skills

Query: "Close Facebook Account"

Time: 2994.5s.

Cumulative Difficulty: 116.5.

Deletion Confirmed

AIFS0020

While these nuances are interesting, the most compelling consequence of this approach is that it allows an assessment of the Equal Clicks Principle from many different perspectives (personas). In figures 5 and 6, each arrow repreents a click. A dashed arrow is an incorrect click (prompting back-tracking), while a blue arrow is a repeated click. Each persona requires 6 clicks to create their account, but HDS requires 19 clicks to delete, ADS 20 clicks and LDS 40. Thus, even accounting for different skill levels, one can be confident that some violation of the ECP is arising in this process. The model also provides an estimation of time. In terms of clicks, deleting an account for a HDS persona is estimated to be around 3 times more effortful than creating an account, and for a LDS persona it is nearly 7 times as effortful.

Using Generated Personas provides additonal insight from a variety of perspectives. Figure 6 shows a replication of a human audit conducted in Mills et al (2023). It provides insight into the LDS persona not discovered by the human auditors. It is also significantly quicker and cheaper.

Conclusion

Generative AI when used as a companion to human auditors can provide cost effective support to protect consumers and vulnerable customers in UK Financial Services whilst supporting the regulator in enforcing the Consumer Duty.

March 2025

AIFS0020

References

Authority, C. M. (2022). Online Choice Architecture: How Digital Design Can Harm Competition and Consumers. In: Discussion Paper.

Authority, F. C. (2022). Finalised Guidance: FG22/5 Final non-Handbook Guidance for firms on the Consumer Duty. In: Guidance Paper, July.

Balasubramnian, B., & Sargent, C. S. (2020). Impact of inflated perceptions of financial literacy on financial decision making. *Journal of Economic Psychology*, *80*, 102306.

Brignull, H. (2011). Dark patterns: Deception vs. honesty in UI design. *Interaction Design, Usability*, *338*, 2-4.

Brown, G., & Whittle, R. (2020). Bitcoin's Threat to the Global Financial System Is Probably at an End'. *The Conversation*, *6*(1).

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of personality and social psychology*, *77*(6), 1121.

Mills, S. (2024a). Being good and doing good in behavioral policymaking. *Public Administration Review*.

Mills, S. (2024b). Deceptive choice architecture and behavioral audits: A principles-based approach. *Regulation & Governance*.

Mills, S., & Whittle, R. (2023). Detecting Dark Patterns Using Generative AI: Some Preliminary Results. *Available at SSRN*.

Mills, S., Whittle, R., Ahmed, R., Walsh, T., & Wessel, M. (2023). Dark patterns and sludge audits: an integrated approach. *Behavioural Public Policy*.

Mills, S., Whittle, R., & Brown, G. (2021). The implications of a crisis-driven societal shift to online consumption. In *Productivity and the Pandemic*. Edward Elgar Publishing.

Mills, S., Whittle, R., & Brown, G. (2024). The Behavioural Dynamics of Social Retail Traders. *Available at SSRN 4831157*.

Newall, P. W. S., & Weiss-Cohen, L. (2022). The gamblification of investing: How a new generation of investors is being born to lose. *International Journal of Environmental Research and Public Health*, *19*(9), 5391.

Packin, N. G., Kliger, D., Reichman, A., & Rabinovitz, S. (2024). Hooked and Hustled: The Predatory Allure of Gamblified Finance. *Available at SSRN 5090194*.

Paez, K. A., & Mallery, C. J. (2014). A little knowledge is a risky thing: Wide gap in what people think they know about health insurance and what they actually know. *American Institutes for Research Issue Brief*, 1-6.

Pompian, M. M. (2012). *Behavioral finance and wealth management: how to build investment strategies that account for investor biases* (Vol. 667). John Wiley & Sons.

Simister, J., & Whittle, R. (2013). Ethical investment and portfolio theory: Using factor analysis to select a portfolio. *Journal of Mathematical Finance*, *3*(1), 145-152.

Whittle, R. (2024). The Consumer Duty: Behavioural Context and Practical Strategies to support informed decision making. Available at SSRN 4810846.

Whittle, R., & Mills, S. (2025). Hard-to-cancel subscriptions: A preliminary exploration of using Generative AI as a behavioural auditing sidekick.