DECEPTIVE PATTERNS PRESSURE TEST

ASSESSING THE RISKS THAT DECEPTIVE PATTERNS POSE TO THE CDR

University of South Australia

Australian Research
Centre for Interactive
and Virtual Environments

A Report to the Data Standards Chair

Pressure Test: Deceptive Patterns

A Report to the Data Standards Chair

James Baumeister
Ji-Young Park
Andrew Cunningham
G. Stewart Von Itzstein
Ian Gwilt
Aaron Davis
James Walsh

Acknowledgements
Graphic Design: Benjamin Altieri

About UniSA's Australian Research Centre for Interactive and Virtual Environments

The University of South Australia's Australian Research Centre for Interactive and Virtual Environments (IVE) is a unique alignment of computer science, engineering, psychology and cognition, neuroscience, art, architecture, and design. Founded in 2019 as a unification of a number of individual areas of expertise, the Centre explores multidisciplinary problems, where the human is at the centre of the solution. The Centre is inspired by the challenges of industry and society to achieve impactful outcomes through delivering world-leading research, developing global research talent, and top-performing PhD students. In collaboration with our industry partners, IVE investigates and combines world expertise in all digital and virtual environments, with computer science, engineering, psychology, neuroscience, art, architecture, and design to solve real-world problems.

Increasingly the problems being encountered in our digital lives are no longer solely technical problems, but problems that touch at the heart of human cognition, emotion, and basic human responses to stimuli. Good digital systems are no longer just created by software developers, rather holistic teams of software developers paired with designers, psychologists, and neuroscientists. Modern applications and web pages are now designed to leverage the user's biological response to stimulus, feeding people's need to infinitely scroll or engage.

IVE's expertise and contribution lies not just in researching and developing solutions for academic problems and industry, but also providing consultation and advice, offering the capability to generate grounded, evidence-based reports and whitepapers, as well as performing grounded, multi-disciplinary objective research focused on the fundamentals of human factors, and how that impacts our relationship with technology.

Contact **IVECentre@unisa.edu.au**

University of South Australia | Australian Research Centre for Interactive and Virtual Environments

UniSA Capability

The University of South Australia (UniSA) is Australia's University of Enterprise and has extensive experience in working with industry, and Defence. UniSA is the largest university in the state with 35,000 students, 2,900 staff, 220,000 alumni and 2,500 partnerships with global universities, research bodies, organisations and industry.

Internationally, UniSA is ranked within the world's top 50 universities under 50 years old, with a five-star rating for World Universities by QS World University Rankings. UniSA is globally recognized as the number 1 young university in Australia for industry collaborations.

In the Excellence in Research for Australia (ERA) assessment, 100% of UniSA research was rated at or above world-class. The university is ranked Number 1 in Australia for industry research impact and engagement and UniSA Business is ranked in the top 1% worldwide.

UniSA is agile and astute, and recognised internationally for relevance, equity and excellence. UniSA educates and prepares global learners from all backgrounds, instilling professional skills, knowledge and a capacity and drive for lifelong learning. UniSA is committed to excellence: excellence in learning, ongoing improvement and innovation, community building, leading effective organisation and management.

The University of South Australia is meeting future challenges through cutting-edge research and the education of tomorrow's professionals.

Disclaimer

This report is not intended to be read or used by anyone other than the Department of the Treasury.

The University of South Australia (UniSA) prepared this report solely for the Department of the Treasury's use and benefit in accordance with and for the purpose set out in the Order of Service with The Department of the Treasury dated 2 February, 2024. In doing so, UniSA acted exclusively for The Department of the Treasury and considered no-one else's interests.

UniSA accepts no responsibility, duty, or liability:

- to anyone other than the Department of the Treasury in connection with this report, or
- to the Department of the Treasury for the consequences of using or relying on it for a purpose other than that referred to above.

UniSA makes no representation concerning the appropriateness of this report for anyone other than the Department of the Treasury. If anyone other than the Department of the Treasury chooses to use or rely on it, they do so at their own risk.

This disclaimer applies:

- to the maximum extent permitted by law and, without limitation, to liability arising in negligence or under statute; and
- even if we consent to anyone other than the Department of the Treasury receiving or using this report, including publication.

Purpose Statement

This report was commissioned pursuant to an Order of Work between the University of South Australia and Department of the Treasury dated 2 February 2024. This report is specifically tailored to the requirements of the Data Standards Chair (Chair) and is to be read within the context of the Consumer Data Right (CDR).

Intended audience

The Chair is the primary owner and audience of this report. The report is also intended to be published and shared with external stakeholders as part of the Chair's requirements to consult.

Executive Summary

This report serves as a pressure test, with the aim of measuring the effectiveness of the Consumer Data Right (**CDR**) at protecting both providers and consumers against the negative influence of Deceptive Patterns (aka Dark Patterns). In the process of our research, we identified 157 Deceptive Patterns within the framework of the IVE Deceptive Patterns Typology. In this report, we analysed each of these deceptive patterns to ascertain whether and how they have the potential to evade the various requirements in the CDR.

Our methodology encompassed two main tests. These tests were formulated as research questions. For the first test, we posed the question as to whether each individual Deceptive Pattern has one or more protective requirements within the CDR's ambit. To provide a concrete answer to this question, we engineered a GPT-4[1] script. This script was specifically designed to compare our definitions of Deceptive Patterns against all 912 active CDR Data Standards, Rules, and Guidelines sourced from their various different origins. The intent was to examine whether Deceptive Patterns were protected against by the CDR Data Standards alone and if not, by other Rules and Guidelines in the overall CDR framework.

The analysis we conducted revealed some interesting insights. We discovered that some Deceptive Patterns, especially those that fall under the Information Asymmetry category of our model, are either completely unprotected or only protected by an optional guideline. This led us to a discussion on the positives and limitations of using Generative Artificial Intelligence (**GAI**) in this manner, which we have included in this report.

For the second test, we undertook an examination of the overall CDR workflows. The aim was to identify if more subtle and creative uses of Deceptive Patterns had the potential to negatively influence any of the processes within these workflows. As a result of this examination, we identified six major concerns. For each of these concerns, we have provided a detailed overview of how the concern can be instigated, what the potential harms are, and we have proposed recommendations for how these concerns could be alleviated.

We hope that this report can serve to bolster the protections of the CDR, and by extension, improve the security and trust of both providers and consumers.

[1] GPT-4 (Generative Pre-trained Transformer 4) is OpenAI's fourth version of its large language model.

Contents

Introduction

This report is an extension of the accompanying landscape assessment entitled **"Patterns in the Dark: Deceptive Practices in Online Interactions"**. In the previous report, we explored the concept of Deceptive Patterns[2], discussing their nature, how they exploit our cognitive vulnerabilities, the different types that exist, and how academic research is identifying the areas where they are having an impact. This report extends that work, focusing specifically on how deceptive patterns impact the CDR.

We define deceptive patterns as follows:

> Deceptive patterns on online interfaces of online platforms are practices that materially distort or impair, either on purpose or in effect, the ability of recipients of the service to make autonomous and informed choices or decisions. Those practices can be used to persuade the recipients of the service to engage in unwanted behaviours or into undesired decisions which have negative consequences for them[3].

This definition is used across both reports.

The CDR is a tool that empowers consumers by giving them more control over their data. It allows them to access and share their data with accredited third parties, enabling them to get better deals on everyday products and services. Key points about the CDR include: it is an opt-in service, where individuals choose whether to share their data, with full visibility of who receives it and for what purpose; it offers benefits, as individuals can compare products, access better services, and manage their finances more easily; the data transfer occurs between providers, is overseen by the Australian Government and co-regulated by the ACCC, and the OAIC; and it is currently implemented in the banking and energy sectors, helping consumers find better products.

We have identified some particular areas of concern regarding deceptive patterns in relation to the CDR. These areas include: **influencing consent, irresponsible and inappropriate use of personal data, hyper-personalised targeting and manipulation, and violation of consumer autonomy and trust.** This report aims to demonstrate how each of these categories of concern can impact some portion of the CDR workflow.

We shall illustrate these points with an example. The following user story is taken from the CDR website[4]:

> **Sarah is a savvy shopper who often looks around for the best interest rate deals on her savings, investment and loan accounts.**
>
> **She also owns a few credit cards from different banks so she can access the best rewards point schemes in the market. She uses her credit cards to pay for many of her everyday expenses, such as shopping, bills and petrol. But Sarah is starting to find all these accounts hard to manage.**
>
> **After some research, Sarah finds the app Consolidata. Reading more on its website, she learns that Consolidata can combine her banking data from all accounts across the different banks in the one place. Confident this will help her manage her finances, Sarah signs on to use Consolidata's service and consents to the collection of her banking data from each of her banks using the Consumer Data Right.**
>
> **Sarah now manages all her accounts in the Consolidata app. While exploring the app, Sarah finds another handy function that alerts her when monthly payments are due, and automatically groups and sorts all transactions so she has a clear picture of her spending habits.**

With Sarah's story in mind, imagine that you are a company in the CDR workflow, either a data holder or a data recipient. From the story, you can make a few observations:

- Sarah will need to provide consent for Consolidata to access her bank information, and the way you present the consent screens to her could influence her behaviour (**influencing consent**).
- Sarah is someone who looks for great deals, a piece of information that you can keep in mind if you want to personalise your approach to Sarah (**hyper-personalised targeting**).
- Sarah uses the same device to access Consolidata as she uses to access many other shopping and social media services, which leave a trace in the form of cookies that you could access and use to build your knowledge base about Sarah's interests (**irresponsible and inappropriate use of personal data**).
- If you can get Sarah's approval to use all this data for direct marketing, you could try to sell her all kinds of other services and pass her data to other companies without her knowledge (**violation of consumer autonomy and trust**).

In the following sections of this report, we will illustrate how Deceptive Patterns can undermine the CDR, why this is problematic, and suggest methods of mitigating these issues.

[2] Deceptive patterns are more commonly referred to as "dark" patterns. In recognition that the usage of "dark" in this way is non-inclusive, UniSA prefers deceptive patterns, which is also a more descriptive term.

[3] (2022) Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance)

[4] https://www.cdr.gov.au/resources/consumer-data-right-stories

Method

The CDR is a collection of consumer experience (**CX**), security, and technical standards. These standards interplay with various sources, including the Competition and Consumer Act (**CCA**) 2010, Competition and Consumer Rules (**Rules**) 2020, and CDR's CX Guidelines. The Data Standards Body (**DSB**) has assembled these requirements into a database[5].

The CDR obligations comprise a range of different types of requirements, such as CX Data Standards, CDR Rules, CX Guidelines, and Technical Data Standards. For the purpose of this report, these will all be referred to as tenets. Altogether, there are 912 active tenets. These tenets are split into requirement levels, one containing mandatory tenets such as "MUST" and "MUST NOT", and the other containing optional tenets such as "MAY" and "SHOULD". It is important to note that the source of the tenet also dictates the obligation of data holders and ADRs to adhere to the tenet, with the Data Standards and Rules representing a mandatory obligation, and Guidelines representing suggestions and best practices, not obligations. The number of tenets per requirement level are as follows:

MUST = 440

MUST NOT = 30

SHOULD = 4

MAY = 438.

The primary focus of this report is to determine whether these standards effectively prevent the negative influence of Deceptive Patterns. The IVE Deceptive Patterns Typology, detailed in the first report and included in Appendix A in this report, consists of 157 deceptive patterns collected from numerous academic sources.

The first test of this report is:

Test 1: Do the CDR standards protect the consumer against each Deceptive Pattern?

To adequately test this, each Data Standard would need to be compared against each Deceptive Pattern. However, this is not feasible for human researchers within the scope of this project. Therefore, a method was developed to incorporate generative AI into the research process.

In order to semantically test whether the CDR standards protect against each Deceptive Pattern, as defined by the IVE Deceptive Patterns Typology (see Appendix A), we employed OpenAI's GPT-4 GAI large language model (**LLM**). We engineered a Python[6] script (see Appendix B) to take each of the 157 deceptive pattern definitions and compare them in turn to each of the 912 Data Standards. The prompt (see highlight) asked the language model to give a 'yes' or 'no' response to the question of whether or not it believed that the statement protected against the concern highlighted by the definition. The curly braced statements (e.g. {deceptive_pattern_name}) represent variables that are exchanged by the script for an actual value (e.g. Roach Motel). Aside from those relating to Deceptive Patterns the others are properties of the standards[7].

This strategy resulted in a total of 157 (deceptive patterns) * 912 (standards) = 143,184 GPT-4 queries. This would have taken an inordinate amount of time for a human, or even a team of humans to do manually. Our script collated all these responses, resulting in a collection of 'yes' or 'no' results to each of the statements[8].

GPT-4 Prompt

You are a helpful assistant.

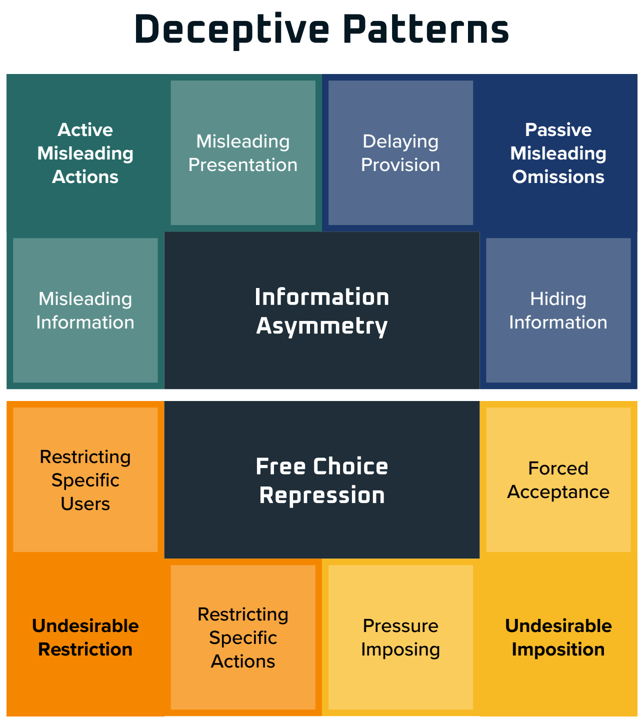
Given the consumer experience rule: {statement} in the focus area: {focus_area} with requirement: {requirement} that participants: {participant} follow, can this consumer rule address the dark pattern[9] named: {deceptive_pattern_name} which has characteristic of {pattern_definition} and reduce the risk associated with the dark pattern: {deceptive_pattern_name}? Respond only with Yes or No. If the consumer experience rule cannot address any dark pattern, you must respond with No.

The second focus of the report is the general insights gleaned from the landscape assessment. The second test is:

Test 2: At what point in the CDR workflow are consumers vulnerable to Deceptive Patterns?

To answer this, we referred to the model presented in the landscape assessment report. By following the various wireframe workflows presented by the CX Guidelines, we identified potential points where deceptive patterns could inject deception and manipulation, negatively impacting the consumer.

Figure 1. Leiser Deceptive Pattern categorisation model. At level 1, patterns are split into either information asymmetry or free choice repression. The four level 2 categories are shown in the corners, with the eight level 3 categories attached.



[5] The CX standards are accessible at https://cx.cds.gov.au/overview/cx-checklist (accessed 07/05/2024).

[6] Python is a general-purpose programming language.

[7] The CX standards are accessible at https://cx.cds.gov.au/overview/cx-checklist (accessed 07/05/2024).

[8] Results available at: https://docs.google.com/spreadsheets/d/1i92BOm1jQPSrTcqpwJLuVpaPYOhPH75NuDld9RlWiMY

[9] We used the more common phrase "dark pattern" here to give GPT-4 broader context.

Results

This section will present the results of our investigation into tests 1 and 2 as outlined in the previous section.

Test 1

Do the CDR standards protect the consumer against each Deceptive Pattern?

According to the IVE Deceptive Patterns Typology (see Figure 1), there are 65 Deceptive Patterns categorised under Information Asymmetry, and 92 Deceptive Patterns fall under the Free Choice Repression category. GPT-4's analysis reveals that the risks associated with 58 deceptive patterns in Information Asymmetry and 89 deceptive patterns in Free Choice can be mitigated by at least one CDR standard. Notably, Deceptive Patterns in Information Asymmetry pose a higher risk based on the current CDR standards.

GPT-4's analysis revealed that 10 Deceptive Patterns (see Table 1) had no protective standards. One interesting Deceptive Pattern to consider is Disgracing Others, defined as "The user is falsely led to believe that a competitor's product is of lesser quality." As will be later discussed, two concerns, which we have named "Notifying the Holder" and "Supplementing the Profile", could use this Deceptive Pattern to discourage consent from the consumer (if used by the data holder), or encourage the consumer to use a non-accredited data recipient (if used by the accredited data recipient (**ADR**)).

Similarly, Fake Exclusive Pricing and Hidden Costs, which both involve convincing a consumer to purchase based on false or misleading pricing, could be used by an ADR to encourage consumers to make a choice for a service that the ADR receives a commission for recommending. This behaviour would be against Australian consumer law, but it is interesting to note that GPT-4 did not find anything explicitly against them in the CDR standards.

Most of the remaining patterns are expectedly irrelevant to the application of CDR. For example, Inducing Artificial Emotions is limited to extended reality devices (such as virtual reality), and Autoplay and Pull-to-refresh are usually found in social media and entertainment applications. Some other patterns, such as Forced Wholesale, Intermediate Currency, and Low-Stock Messages relate to e-commerce, which is also outside CDR purview.

Deceptive Pattern	May	Must	Must Not	Should	Total	Relevant
Autoplay	0	0	0	0	0	X
Disgracing Others	0	0	0	0	0	✓
Display Controversial Content	0	0	0	0	0	X
Fake Exclusive Pricing	0	0	0	0	0	✓
Forced Wholesale	0	0	0	0	0	X
Hidden Costs	0	0	0	0	0	✓
Inducing Artificial Emotions	0	0	0	0	0	X
Intermediate Currency	0	0	0	0	0	X
Low-stock Messages	0	0	0	0	0	X
Pull-to-refresh	0	0	0	0	0	X

Table 1. All Deceptive Patterns with zero protective CDR standards.

Table 2 shows the 14 Deceptive Patterns that GPT-4 found to be only protected by one or more optional tenets; those marked as having a 'May' requirement level. As before, some of these patterns can be immediately marked as irrelevant to CDR's operational environment. Of the 14, six of them were deemed relevant. Five of those (Fake Scarcity, Fake Social Proof, Testimonials of Uncertain Origin, High Demand Messages, and Scarcity) reside within the Misleading Information category of the IVE Deceptive Patterns Typology (see Figure 1).

These five patterns all represent deceptive sales tactics that commercial entities can use to either promote a sense of decision-making urgency (Fake Scarcity / Scarcity, High Demand Messages) or instil confidence in a product by use of popularity or peer-based high regard (Fake Social Proof, Testimonials of Uncertain Origin). Our concern with these patterns is not that the consumer's data or privacy is violated, but that trust in an ADR, and the CDR by extension, could be compromised if an ADR were to employ these types of deception.

Consider a hypothetical scenario related to the previous example of Sarah using the Consolidata app. According to GPT-4's findings, Consolidata could encourage Sarah to purchase an additional service that can help her better understand her budget and improve her ability to save money. This sounds great, but the app could promote this service by including statements such as, "people who have similar spending habits to you tend to save more money when they add our savings maximiser package". Using the Fake Social Proof Deceptive Pattern, Consolidata may manipulate Sarah into purchasing this package due to an unsubstantiated claim. This is just an example; we do not claim that any ADR is currently doing this.

The remaining relevant Deceptive Pattern is Nickling-and-diming, defined as "the user is prevented from interacting with a service by an initially disguised requirement for payment." As with the other examples, this Deceptive Pattern could be employed outside the regular CDR authentication and consent workflows.

We can imagine two scenarios where this pattern could be applied. In the first, an ADR could encourage a consumer to link their banking or energy services and once that occurs, the most useful aspects of the ADR's app could be withheld behind a paywall. After having expended some effort into connecting services, consumers may be more likely to accede to the payment. Even if they do not, the linking process has given the ADR some information that may be valuable to them. In the second scenario, relating specifically to ADRs in the finance sector, the ADR could use the consumer's finance data to suggest that they spend more money, perhaps with one of the ADR's retail partners. They could apply other incentives, such as discounts or coupons, to encourage spending from which the ADR would receive a commission.

Examining the types of statements that are found to be protective in Table 2, it is worth noting that a select few, very broad statements (1CO1.01.31, 5CM1.00.16, 4CM1.00.25, 5CM1.00.20, 1CO.00.37, 1CO1.02.08, 1CO2.02.30, 1CO4.00.33) are doing the protecting. In fact, 1CO1.01.31, 1CO.00.37, and 1CO1.02.08 are identical, with 5CM1.00.16 being a slightly shortened version. Similarly, 4CM1.00.25 and 5CM1.00.20 are also identical. This means that there, in effect, only four CX Guidelines with an optional requirement protecting against six relevant Deceptive Patterns.

We propose that none of these statements specifically address the highlighted Deceptive Patterns and that GPT-4's decision to mark them as protective is based on very broad protection from phrases such as "easy to understand" and "reduces cognitive overload". There is much room for interpretation here.

Table 2. All deceptive patterns with only optional protective CDR standards. Included are the statements that GPT-4 believed are protective, along with their CX Checklist ID. All statements are from the CX Guidelines.

Deceptive Pattern	May	Must	Must Not	Should	Total	Relevant
Fake Scarcity	1	0	0	0	1	✓
Data recipients should make the consent process as easy to understand as possible. Data recipients should nudge consumers to be more privacy conscious and should use appropriate interventions to mitigate cognitive overload, facilitate comprehension, and provide transparency and consumer control. This can be done in a variety of ways, including through the use of design patterns like progressive disclosure, micro and/or descriptive copy, and with the use of microinteractions.						**1CO1.01.31 CX Guideline**
Fake Social Proof	1	0	0	0	1	✓
As above						**1CO1.01.31 CX Guideline**
Infinite Scrolling	1	0	0	0	1	X
If scrolling is required to view the total number of CDR participants, data holders should provide search functionality.						**5CM1.00.10 CX Guideline**
Misleading Experience Marketing	1	0	0	0	1	X
As above						**CO1.01.31 CX Guideline**
Nickling-and-diming	1	0	0	0	1	✓
As above						**CO1.01.31 CX Guideline**
Playacting	1	0	0	0	1	X
Data recipients should seek to, for example, describe data concisely, in plain language, with an Australian year 7 or lower readability level, and in a way that limits the use of unusual words, phrases, idioms, and jargon.						**1CO3.00.23 CX Guideline**
Social Pyramid	1	0	0	0	1	X
Data holders should nudge consumers to be more privacy conscious and should use appropriate interventions to facilitate comprehension and consumer control. This can be done in a variety of ways, including through the use of design patterns like progressive disclosure, micro and/or descriptive copy, and with the use of micro-interactions.						**5CM1.00.16 CX Guideline**
Testimonials of Uncertain Origin	1	0	0	0	1	✓
As above						**5CM1.00.16 CX Guideline**
Social Investment	2	0	0	0	2	X
As above						**1CO1.01.31 CX Guideline**
Spoof Content	2	0	0	0	2	X
As above						**1CO1.01.31 CX Guideline**
Data recipients may meet standards requirements in relation to non-accredited person data handling at appropriate points throughout the Consent Model, such as: during Pre-consent; during Consent, as required by the data standards in relation to data handling and disclosure notifications; within the CDR Receipt and/or Consumer Dashboards, as required by the disclosure notification standards.						**1CO3.01.19 CX Guideline**

Deceptive Pattern	May	Must	Must Not	Should	Total	Relevant
Gamification	3	0	0	0	3	X
As above						1CO1.01.31 CX Guideline
To describe data in easy to understand language, data recipients should have regard to the Accessibility Standards on reading experiences, with specific reference to WCAG 3.1.5, and draw from the Australian Government Style Manual on literacy and access. Data recipients should seek to, for example, describe data concisely, in plain language, with an Australian year 7 or lower readability level, and in a way that limits the use of unusual words, phrases, idioms, and jargon.						1CO4.00.33 CX Guideline
As above						5CM1.00.16 CX Guideline
High-demand Messages	4	0	0	0	4	✓
As above						1CO1.01.31 CX Guideline
Data recipients should prioritise information that is important to consumers and structure the presentation in a way that reduces cognitive overload. This may include progressive disclosure design patterns (e.g. accordion menus), UX writing (e.g. microcopy), and visual aids (e.g. to display time-based qualities of consent).						4CM1.00.25 CX Guideline
As above						5CM1.00.16 CX Guideline
Data holders should prioritise information that is important to consumers and structure the presentation in a way that reduces cognitive overload. This may include progressive disclosure design patterns (e.g. accordion menus), UX writing (e.g. microcopy), and visual aids (e.g. to display time-based qualities of consent).						5CM1.00.20 CX Guideline
Ad Drop-down Delay	6	0	0	0	6	X
As above						1CO1.01.31 CX Guideline
Data recipients should make the consent process as easy to understand as possible. Data recipients should nudge consumers to be more privacy conscious and should use appropriate interventions to mitigate cognitive overload, facilitate comprehension, and provide transparency and consumer control. This can be done in a variety of ways, including through the use of design patterns like progressive disclosure, micro and/or descriptive copy, and with the use of microinteractions.						1CO1.02.08 CX Guideline
As above						4CM1.00.25 CX Guideline
Data recipients should prioritise information that is important to consumers and structure the presentation in a way that reduces cognitive overload. This may include progressive disclosure design patterns (e.g. accordion menus), UX writing (e.g. microcopy), and visual aids (e.g. to display time-based qualities of consent).						4CM1.01.23 CX Guideline
As above						5CM1.00.16 CX Guideline
As above						5CM1.00.20 CX Guideline

Deceptive Pattern	May	Must	Must Not	Should	Total	Relevant
Scarcity	8	0	0	0	8	✓
Data recipients should make the consent process as easy to understand as possible. Data recipients should nudge consumers to be more privacy conscious and should use appropriate interventions to mitigate cognitive overload, facilitate comprehension, and provide transparency and consumer control. This can be done in a variety of ways, including through the use of design patterns like progressive disclosure, micro and/or descriptive copy, and with the use of microinteractions.						1CO.00.37 CX Guideline
As above						1CO1.01.31 CX Guideline
As above						1CO1.02.08 CX Guideline
Data recipients should communicate that consent will expire if request is not actioned.						1CO2.02.30 CX Guideline
As above						1CO4.00.33 CX Guideline
As above						4CM1.00.25 CX Guideline
As above						5CM1.00.16 CX Guideline
As above						5CM1.00.20 CX Guideline

Our GPT-4 analysis has shown that the CDR is protective against the vast majority of Deceptive Patterns. As a proof of concept, we have shown that a LLM can be tasked with comparing large datasets comprising complex statements and legal definitions and that the results can be insightful and consistent. GPT-4 found that similar patterns within the Misleading Information category of our deceptive patterns model are unprotected by mandatory standards. By performing an enormous number of comparisons within a short space of time, GPT-4 proved to be a cost-effective and useful tool for reducing the breadth of work that human analysts are required to attend.

AI Limitations

There are several limitations to the GPT-4 approach that we employed that must be acknowledged. Firstly, the LLM that GPT-4 utilises to comprehend the Deceptive Pattern definitions and the tenets does not have all the necessary context that an expert human reviewer would have. For example, many tenets refer to sections or statements in other documents, such as in 0AC.05 (see CX Checklist[10]): “Data recipients and data holders MUST seek to have all aspects of the Consent Model comply with WCAG 3.3. This will help users avoid and correct mistakes.” In this statement, reference is made to WCAG 3.3, which GPT-4 is not told to consult in the prompt. This means that the result may not be accurate without this necessary context. This requirement for additional context and domain knowledge, however, would also apply to an app developer who is trying to abide by the many complex and fragmented requirements. Reference material, such as WCAG, could be added to further developed uses of GAI LLMs for this purpose.

Another limitation pertains to the reasoning behind the language model’s decision of ‘yes’ or ‘no’. We limited the response to this binary decision to limit our search space for further examination, but this leaves no explanation as to why that decision was made. If our prompt had included an expectation for GPT-4 to provide a reason for its choice, it would have given one, but the reason cannot be relied upon. This is due to the explanation being based on the training that went into the development of the LLM, not actually the internal processes that resulted in the decision.

To give a human analogy, if you ask your friend to choose a restaurant for dinner, they will provide one. If you then ask why, your friend may give you a detailed reason, perhaps including proximity, cuisine preference, and cost, but this response is not actually a true representation of the inner, electrical workings of the brain that led to your friend to give you their reasons.

Our results revealed that GPT-4 is effective at comprehending the semantic meaning of CDR standards but that there is no guarantee that ‘yes’ or ‘no’ responses are entirely accurate. Our methodology is a proof of concept and GPT-4 can be a useful tool to accelerate otherwise tedious and manual processes, but not reliable in isolation. Future research could investigate how the prompt can be further engineered to produce more effective outcomes, and also how extra context (for example, legislation and other reports) could be provided to GPT-4 or another LLM, such as Microsoft Copilot, to give it extra information with which to make a determination. Additionally, future research could test how fine-tuning the model with training data focussed on CDR-related materials can impact the consistency and explainability of the results.

[10] The CX Checklist is accessible at https://cx.cds.gov.au/overview/cx-checklist (accessed 07/05/2024)

Test 2

At what point in the CDR workflow are consumers vulnerable to deceptive patterns?

Test 1 showed that the GPT-4 could identify some unprotected deceptive patterns. However, GPT-4 was quite literal in comparing one standard against one deceptive pattern. This section represents our effort to consider the overall context of the CDR and the IVE Deceptive Patterns Typology. We present six concerns in this section, along with a description of why they present a potential problem for the CDR and give examples of how the concern could be instantiated within the CDR workflow.

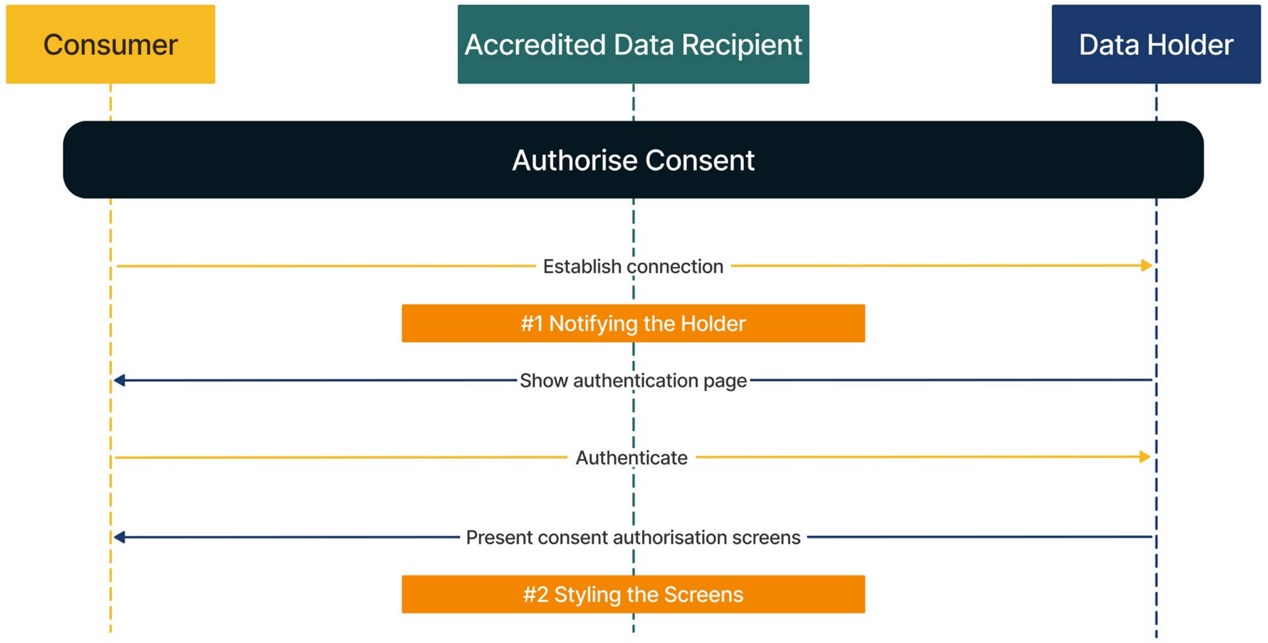
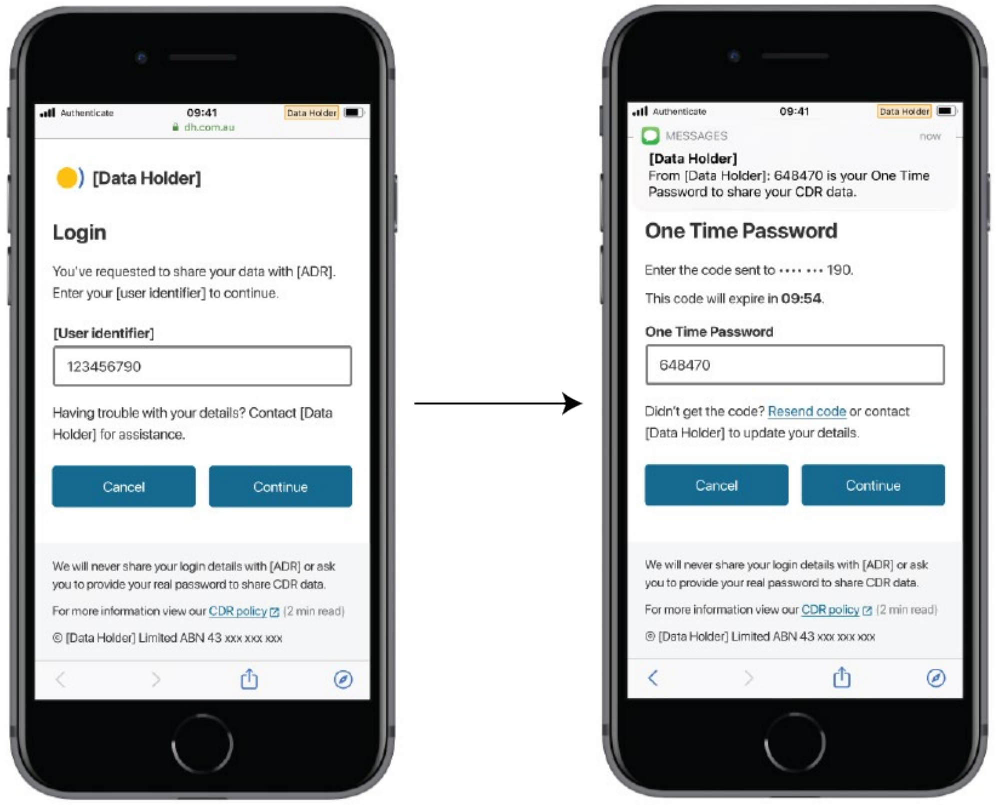


Figure 2. A sequence diagram representation of the simplified CDR consent process. Shown in orange are the potential problems #1 and #2.

#1 Notifying the Holder

Related Deceptive Patterns: Inducements to Reconsider | Retaining Customers | Last Minute Solutions



[Data Holder]

Login

You've requested to share your data with [ADR]. Enter your [user identifier] to continue.

[User identifier]: 123456790

Having trouble with your details? Contact [Data Holder] for assistance.

We will never share your login details with [ADR] or ask you to provide your real password to share CDR data.

For more information view our CDR policy (2 min read)

© [Data Holder] Limited ABN 43 xxx xxx xxx

[Data Holder]
From [Data Holder]: 648470 is your One Time Password to share your CDR data.

One Time Password

Enter the code sent to •••• ••• 190.

This code will expire in **09:54**.

One Time Password: 648470

Didn't get the code? Resend code or contact [Data Holder] to update your details.

We will never share your login details with [ADR] or ask you to provide your real password to share CDR data.

For more information view our CDR policy (2 min read)

© [Data Holder] Limited ABN 43 xxx xxx xxx

Figure 3. A prototype user interface for the hand-off between the ADR and data holder during the authentication process.

When a consumer initiates a data share with the ADR, the data holder becomes aware of this connection. This awareness offers the data holder a significant amount of information about the consumer's interests and intentions. It reveals that the consumer is interested in linking with the data recipient service and it provides specifics about the service they are linking with.

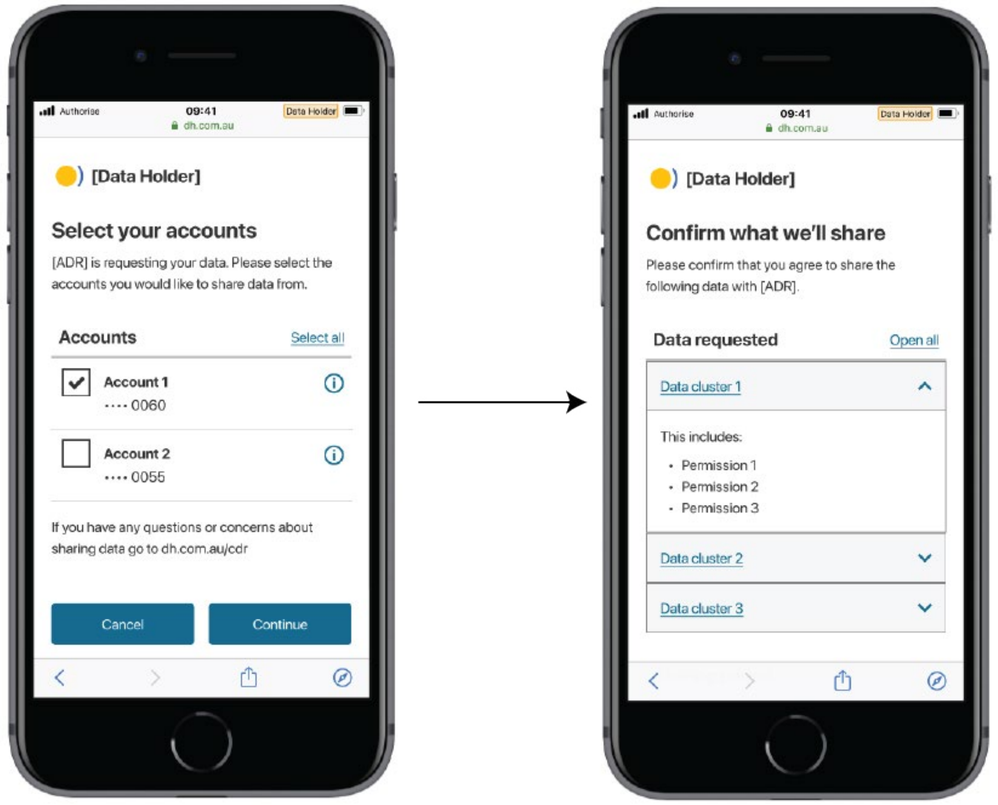
The data holder can utilise this information to modify their interaction with the consumer outside of CDR-protected workflows. For instance, the data holder could attempt to persuade the consumer to abandon the ADR's service by actively promoting their own similar tools. This action could take place entirely within the data holder's own app or communications. It is important to note that such actions are not governed under CDR, thus providing the data holder with an avenue to influence consumer behaviour without breaching the standards.

Considerations

It is recommended to consider placing restrictions on metadata storage for incomplete or cancelled consent flows. If the consumer does not complete the consent workflow, any associated metadata should be discarded. This protects the consumer from any targeting or changes in data holder or ADR services as a result of incomplete CDR links. It is also recommended to specify how authorisation and consent metadata can be used for successful links.

#2 Styling the Screens

Related Deceptive Patterns: Asymmetric Button | Bad Visibility | Chameleon Strategy | Colour | Visual Interference



[Data Holder]

Select your accounts

[ADR] is requesting your data. Please select the accounts you would like to share data from.

☑ Account 1 •••• 0060

□ Account 2 •••• 0055

If you have any questions or concerns about sharing data go to dh.com.au/cdr

[Data Holder]

Confirm what we'll share

Please confirm that you agree to share the following data with [ADR].

This includes:

- Permission 1
- Permission 2
- Permission 3

Figure 4. A prototype user interface for an authorise screen that a data holder can present to request consumer consent to link the holder to the ADR.

When the ADR transitions to the data holder for authorisation, the styling of these screens is largely determined by the data holder. The holder is mandated by the CDR to provide information about what is being shared, why, and for how long. The holder must then offer the consumer the opportunity to either consent to the sharing of their information with the ADR or decline the share.

If the data holder has no motivation to influence the consumer towards a particular decision (consent or decline), the CDR clearly stipulates that information be clearly presented and the choice given to the consumer. However, different motivations may exist. For instance, a consumer might want to explore various energy plans and find an ADR app that can analyse their current electricity usage and present alternative, cheaper options.

When the consumer requests the ADR to link their existing electricity accounts to share their usage data, the holder is given some information about their customer shopping around for a better deal. It is in the holder's commercial interest to retain the customer. Therefore, it is plausible that the holder may attempt to reduce the likelihood that the customer establishes the CDR connection and stays with their existing services.

In the 'Styling the Screens' concern, there are very subtle visual user interface manipulations that holders can employ that are not explicitly in violation of CDR.

An example is the Asymmetric Button deceptive pattern. On the authorisation screen, the data holder could visually emphasise, perhaps with colour, size or placement, the decline button rather than the consent button.

Another Deceptive Pattern that could aid this is Visual Interference. The CDR requires the authorisation screen to contain a lot of information. If the information is presented in such a way that the consumer is overwhelmed, then when they reach their options (consent or decline) and one is more prominent, they may choose to decline more often.

As discussed in the landscape assessment, Artificial Intelligence-powered A/B testing, where various interfaces are presented until the most effective interface is identified, could help a data holder discover which CDR-compliant Deceptive Patterns best aid their aim to influence the consumer's consent choice.

Both data holders and ADRs could employ several Deceptive Patterns related to information presentation. These patterns are typically extremely subtle, and likely only influence the choice of a small number of consumers. Despite their subtlety, if even a small percentage of consumers are more likely to perform a behaviour that aligns with a commercial entity's agenda, this could prove profitable. One pattern identified was the Asymmetric Button pattern, where multiple buttons are designed with one appearing more prominent or appealing than the others. In the context of the CDR workflow, data holders or ADRs could subtly style the cancel or consent button, depending on their preferred outcome for the consumer. For instance, the consent button could be made slightly smaller, with duller text if the preferred action is for the consumer to push the cancel button.

Other patterns such as the Bad Visibility, Chameleon Strategy and Visual Interference are all used to achieve the same goal of visually disguising user interface (UI) components that they would prefer the consumer to miss. A common example of this is hiding a checkbox that enables a consumer to opt out of email communication inside a collapsed UI element. The Colour Deceptive Pattern can be especially difficult to detect and regulate, as the use of colour is very important for effective communication and user experience. When combined with the aforementioned Asymmetric Button pattern, colour could be used to design the preferred button in the data holder or ADR's corporate branding, leaving the non-preferred option in a dull colour. This could subtly indicate to the consumer that the preferred option is trusted, as they may already trust the colours associated with the brand.

Considerations

It is suggested to create mandatory webview templates for consent workflows, both for data holders and ADRs. These can include customisable components for corporate branding. This will eliminate style choices with subtle deceptive patterns. Specific requirements for presenting webviews should also be listed, such as no third-party tracking libraries (for example, Hotjar[11]), sanitisation of cookies, and no handing off the mandatory webview with a redirect.

[11] https://www.hotjar.com/

#3 Supplementing the Profile

Related Deceptive Patterns: Shadow User Profiles | Unintended Relationships

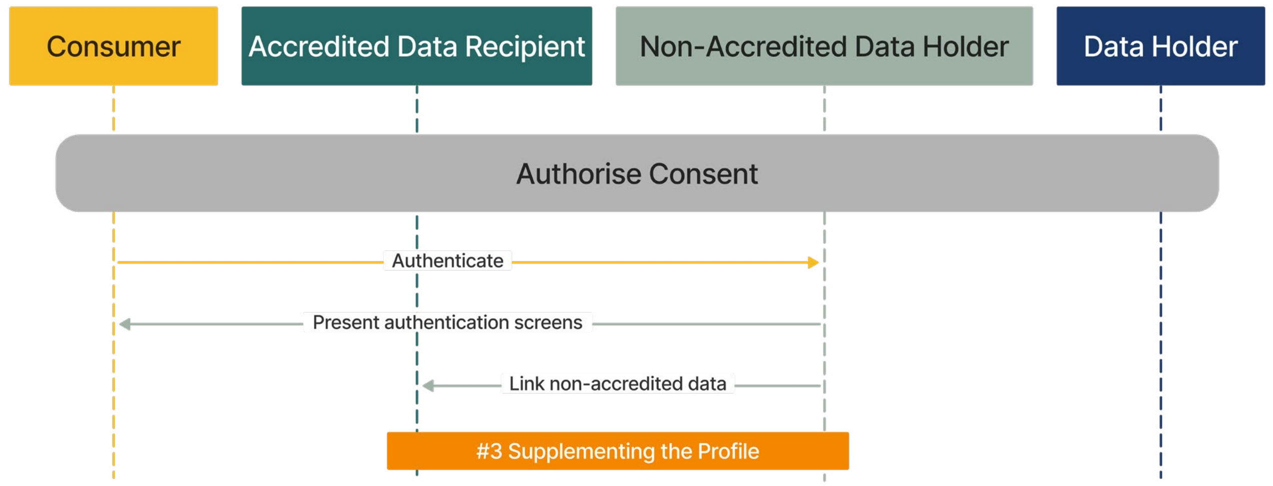

Consumer
Accredited Data Recipient
Non-Accredited Data Holder
Data Holder
Authorise Consent
Authenticate
Present authentication screens
Link non-accredited data
#3 Supplementing the Profile

Figure 5. A sequence diagram representation of a consent process between an ADR and a non-ADR. Show in orange is is potential problem #3.

ADR services have the ability to connect to other apps and services that are not members of the CDR. This connection provides the data recipient with an additional layer of information about the consumer that would ordinarily be prohibited by the CDR. This additional data can be used to construct a complex consumer profile from multiple sources. In most cases, this is a desirable behaviour for the consumer as it enables useful insights to be gleaned from the aggregation of multiple data sources.

For example, trends such as "50% of service users spend more than they are making" can inform users about whether they are in certain groups and can action that information. The concern, however, is how to detach CDR-derived data from other data sources when used for aggregations such as these.

The fact that these sources include those not governed by the CDR makes the provenance of the data opaque, giving the data recipient more leeway to misuse the information. In certain scenarios, the data recipient could potentially link CDR and non-CDR data sources to build a detailed consumer profile that the consumer themselves is not aware of. This raises significant concerns about consumer privacy and data protection.

Considerations

The main consideration is that many Deceptive Patterns, particularly those within the Free Choice Repression categories (see Figure 1) rely on having a database of information with which commercial entities can focus the Deceptive Pattern toward a consumer and improve its efficacy. Although the Supplementing the Profile concern contains data that is outside the CDR workflows, it is possible that CDR and non-CDR data can be combined to power highly effective, targeted Deceptive Patterns. It is worth considering whether existing protections of CDR data are protective enough to prevent the kinds of data misuse described in this concern, including seemingly benign metadata. This applies to when it is used with other non-CDR sources. Examples of how combining CDR and non-CDR data can be beneficial to the consumer should be provided. Wherever CDR data has been combined or aggregated with non-CDR data, however, it cannot be sold to a data-broker or used in any way not in compliance with CDR.

#4 Anticipating the Renewal

Related Deceptive Patterns: Continued Email Communication | Rewards and Punishment | Safety Blackmail

When a consumer creates a connection between a data holder and an ADR, similar to the 'Notifying the Holder' section, the holder becomes aware of this agreement. This awareness extends to the important dates related to the agreement. These include the quarterly data linkage reminders and the agreed upon renewal date.

Having this knowledge in advance, the data holder could potentially use this information to their advantage, to influence the consumer's decision to renew the agreement. For example, ahead of the anticipated renewal date, a data holder could subtly suggest competing products or offer discounts on any paid services. This strategy can encourage the consumer to perceive the renewal as unnecessary and consider switching to the data holder's services.

Also relating to upcoming renewals is how an ADR contacts a consumer to remind them of upcoming consent expirations. As an example, the ADR Frollo sends an email for each linked data holder, reminding the consumer to renew their consent. This is CDR compliant, is useful information for the consumer, and is a desirable behaviour from the CDR. We also know, however, that consumers often receive an overwhelming number of emails and having a large number of them appear simultaneously for basically the same topic (the only difference would be the name of the data holder) could lead to complacency and a tendency to either ignore all of them, or decide upon an action for all data holders after reading only one email.

Considerations

It is crucial to protect the consumer's autonomy from undue influence from either the data holder or ADR. The choice must be presented to the consumer, and their right to choose cannot be influenced by the use of deceptive patterns, even outside CDR workflows. Renewal notices should be encouraged to be batched to avoid email overload.

#5 Middleman the Requests

Related Deceptive Patterns: Recommendations | Hyperpersonalisation | Unintended Relationships

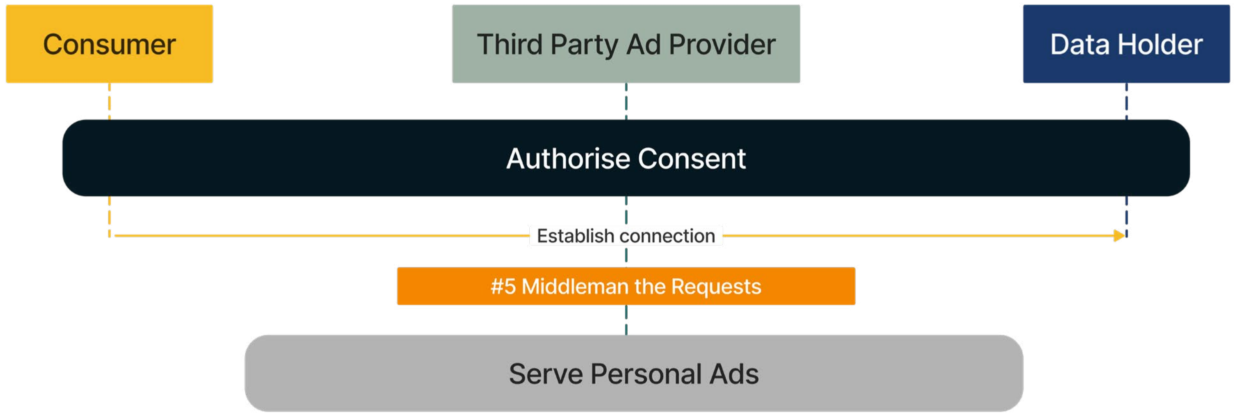


Figure 6. A sequence diagram representation of the simplified CDR consent process and the interception of remaining meta-data from that process. Shown in orange is the potential problem #5.

The methods of authentication and authorisation utilised by data holders can vary significantly. For some data holders, these processes may take place within their own app. For others, they may occur in a browser or webview. This distinction is critical as it determines how and where consumer data is stored and accessed.

When authentication and authorisation occur in a browser or webview, it creates a trail of information, typically in the form of cookies, on the consumer's device. These cookies can be detected and utilised by other services, such as ad providers. For instance, an ad provider could detect a cookie created during a CDR authorisation workflow and use this information to target the consumer with relevant advertisements. It is potentially more dangerous than just targeted advertisements, however. Cookies can contain seemingly benign information about a user's location, device IP, session meta-data, and many other data points, but this information can be very valuable to data brokers. If the webview is not sanitised after the authentication and authorisation workflows, information could remain and be accessible to other malicious entities on the user's device.

Another issue arises when some data holders, lacking the necessary software infrastructure within their own apps to handle the CDR-compliant authentication workflows, outsource this task to third parties. This action implies that some CDR data is now in the possession of a third party, making the data holder's role in ensuring compliance more complex and potentially expensive. It also involves placing a significant amount of trust in third parties who might not be accredited under CDR and may have commercial incentives to misuse the data. This bears striking similarity to the issue of commercial cookie management platforms (CMPs), which our other report noted as many being GDPR compliant, yet offer deceptive patterns as part of their service.

Considerations

The same requirements as styling the screens should be applied. Specific requirements for presenting webviews should be listed, such as no third-party tracking libraries, sanitisation of cookies, and no handing off the mandatory webview with a redirect.

#6 Market the Consumer

Personalisation stands as a pivotal trait of sophisticated Deceptive Patterns. One area that this aspect of personalisation is particularly evident is in direct marketing, especially with the use of CDR data. The directives surrounding the use of CDR data for direct marketing are primarily covered under Privacy Safeguard 7[12]. Interestingly, this safeguard is centred around the ADR, not the data holder.

While it prohibits the ADR from using or disclosing CDR data for direct marketing purposes, this safeguard does not apply to the data holder. The data holder, on the other hand, is required to comply with Australian Privacy Principle 7[13]. As per the current regulations of this principle, a data holder is permitted to use and disclose personal information, excluding sensitive information, for direct marketing under certain conditions. In light of the vast amount of data available in the modern world, data providers can construct highly detailed user profiles for their businesses without even needing sensitive personal data. These profiles can then be exploited to generate highly effective personalised Deceptive Patterns, including personalised recommendations.

Considerations

To counteract the potential risks associated with the use of such user profiles, we propose the introduction of user consent in direct marketing scenarios on the data holder's side.

[12] Available at https://www.oaic.gov.au/__data/assets/pdf_file/0014/111560/63e3c41172c371facadae7dda21bc40e2a8cce5f.pdf (accessed 30/05/2024)

[13] Available at https://www.oaic.gov.au/privacy/australian-privacy-principles/read-the-australian-privacy-principles (accessed 30/05/2024)

Conclusion

This report has detailed our approach to pressure testing the CDR against the IVE Deceptive Patterns Typology, an approach which we have divided into two parts. In general, our advice to the Data Standards Chair is to consider how the subtlety of Deceptive Patterns can skirt existing Standards, Rules, and Guidelines to subtly shift consumer behaviour toward the agendas desired by data holders or ADRs. In most cases, the CDR workflows, including all regulatory and non-regulatory sources (which we have termed 'tenets'), protect against the most egregiously nefarious Deceptive Patterns.

The first part of our pressure test, which utilised the GAI LLM GPT-4 to compare our Deceptive Pattern definitions against all active tenets, revealed that some patterns should be closely examined to confirm whether any additional protections are required to prevent the types of negative influence we have outlined. It is worth noting a trend that Deceptive Patterns within the Information Asymmetry category are more likely to be either completely unprotected by any tenet, or only protected by a Guideline, which neither data holders nor ADRs are obliged to follow.

Furthermore, we noticed that only a few, very broadly phrased Guidelines are offering the protection. Given the subtlety with which Deceptive Patterns can be implemented and how difficult they can be to detect, it might be worth considering whether having more specifically phrased protections may be better than relying on tenets that are not specifically designed to protect against them. We believe that our approach to this first pressure test, based on GPT-4, was effective in highlighting some areas of weakness in the overall CDR framework.

The second part of our pressure test involved an observation of the CDR workflows and an evaluation of where Deceptive Patterns might negatively influence the consumer and skirt existing protections. We outlined six concerns and suggested how the Chair might like to consider consulting CDR domain experts on how these concerns might be addressed. Some of the concerns specifically relate to the CDR workflows, either regarding how information is presented to the consumer (#2), or how consumer data could be hijacked in ways not intended by CDR (#1, #4, and #5). These concerns could largely be addressed with some additional directives that specifically prevent the design behaviours that we outline.

There were two other concerns that might appear to be outside the scope and purview of the CDR (#3 and #6), which outline how data or metadata produced by CDR workflows could be collected and used to modify how data holders, ADRs, and potentially non-ADRs, interact with consumers.

As illustrated in our associated report, many of the more dangerous Deceptive Patterns can be powered by AI and hyper-personalised in their targeting of consumers. We suggest that the Chair consider that if these concerns do indeed fall beyond the scope of the CDR workflows, that further investigation be conducted into whether other legislation (e.g. Privacy Act, Competition and Consumer Act, Spam Act, etc) effectively offer the necessary protection. Our research (see associated report) has indicated that Australian legislation is not effective enough at preventing the subtleties of many Deceptive Patterns, and that even international regulatory pieces (e.g. GDPR) are not prepared for the arising threat of AI-powered Deceptive Patterns.

Our report has identified and addressed potential areas of concern in the CDR framework with regards to the potential for Deceptive Patterns. Our recommendations aim to ensure that the CDR remains as robust and protective as possible, keeping the consumer's best interests at its core.

Appendix A

Table 3. IVE Deceptive Patterns Typology

<table>
<tr><th>Name</th><th>Level 1</th><th>Level 2</th><th>Level 3</th><th>Source</th></tr>
<tr><td>Activity Notifications</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Mathur et al., 2019</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is misled into believing a product is more popular or credible than it really is, because they were shown activity messages.</td></tr>
<tr><td>Address Book Leeching</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Bösch et al., 2016</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is prompted to give a service access to their address book to connect with known contacts also on the service, but other purposes are not declared.</td></tr>
<tr><td>Disgracing Others</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Wu et al., 2022</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is falsely led to believe that a competitor’s product is of lesser quality.</td></tr>
<tr><td>Fake</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Bösch et al., 2016</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is presented a “universally” understood graphic code but the meaning is opposite to the expected.</td></tr>
<tr><td>Fake Exclusive Pricing</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Wu et al., 2022</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is convinced to purchase based on a fake, exclusive, or discounted price that was raised before the discounted price was advertised.</td></tr>
<tr><td>Fake Scarcity</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Mathur et al., 2019</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is pressured into completing an action because they are presented with a fake indication of limited supply or popularity.</td></tr>
<tr><td>Fake Social Proof</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Mathur et al., 2019</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is misled into believing a product is more popular or credible than it really is, because they were shown fake reviews, testimonials, or activity messages.</td></tr>
</table>

<table>
<tr><th>Name</th><th>Level 1</th><th>Level 2</th><th>Level 3</th><th>Source</th></tr>
<tr><td>Fake Urgency</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Mathur et al., 2019</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is pressured into completing an action because they are presented with a fake time limitation.</td></tr>
<tr><td>False Necessity</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Kitkowska, 2023</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is falsely informed that certain types of data are legally necessary or required for the system to function.</td></tr>
<tr><td>Framing</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Norwegian Consumer Council, 2018</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is shown information that positively frames the consequences of an action, while omitting the entailed risks.</td></tr>
<tr><td>Hidden Legalese Stipulations</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Bösch et al., 2016</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is misled by complicated legal jargon to accept a legally binding policy without understanding the implications.</td></tr>
<tr><td>High-demand Messages</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Mathur et al., 2019</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is presented a message stating that a product is in high demand, implying that it will likely sell out.</td></tr>
<tr><td>Just Between You and Us</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>National Commission on Informatics and Liberty (CNIL), 2020</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is promised that additionally provided information will remain invisible but ultimately provide a better service.</td></tr>
<tr><td>Lie</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Conti and Sobiesk, 2010</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is presented with an outright lie, such as them winning a contest.</td></tr>
<tr><td>Limited-time Messages</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Mathur et al., 2019</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is presented a message stating that a product is only available for a limited time.</td></tr>
</table>

<table>
<tr><th>Name</th><th>Level 1</th><th>Level 2</th><th>Level 3</th><th>Source</th></tr>
<tr><td>Loss-gain Framing</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Bongard-Blanchy et al., 2021</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is shown information that positively frames the consequences of an action, while omitting the entailed risks.</td></tr>
<tr><td>Low-stock Messages</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Mathur et al., 2019</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is presented a message stating that a product is in low stock, implying that it will likely sell out.</td></tr>
<tr><td>Misrepresenting</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Gray et al., 2020</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is presented ambiguous and incorrect information in order to trick them.</td></tr>
<tr><td>Misunderstood Questions</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Conti and Sobiesk, 2010</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is asked questions that use confusing language, such as double, triple, or quadruple negatives.</td></tr>
<tr><td>Scarcity</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Gray et al., 2023</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is pressured into completing an action because they are presented with a fake indication of limited supply or popularity.</td></tr>
<tr><td>Sophistry</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Wu et al., 2022</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is shown information that positively frames the consequences of an action, while omitting the entailed risks.</td></tr>
<tr><td>Testimonials of Uncertain Origin</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Mathur et al., 2019</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is misled into believing a product is more popular or credible than it really is, because they were shown fake testimonials.</td></tr>
<tr><td>Two-faced</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Information</td><td>Gray et al., 2020</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is shown contradictory and conflicting information.</td></tr>
</table>

Name	Level 1	Level 2	Level 3	Source
Violate	Information Asymmetry	Active Misleading Actions	Misleading Information	Bösch et al., 2016
Definition: The user is presented a privacy policy that is intentionally violated by the presenter.				
Wrong Signal	Information Asymmetry	Active Misleading Actions	Misleading Information	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is presented a "universally" understood graphic code but the meaning is opposite to the expected.				
Asymmetric Button	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Long et al., 2023
Definition: The user is directed by button size and colour to gravitate toward options that do not align with their intentions.				
Bad Visibility	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Kitkowska, 2023
Definition: The user is offered options where desirable options (undesirable to the service) are presented with low contrast, light colours, and small fonts.				
Chameleon Strategy	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Kitkowska, 2023
Definition: The user is presented with a third-party service that mimics the style and visual appearance of the original service to make it look like a natural continuation.				
Colour	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Conti and Sobiesk, 2010
Definition: The user's attention is guided to a designer's preference by attractive colour use.				
Dead End Trails	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Conti and Sobiesk, 2010
Definition: The user is presented by seemingly endless questions ostensibly to result in a desired outcome.				
Distorting Reality	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Mhaidli and Schaub, 2021
Definition: The user is presented, via extended reality (XR) a distorted version of reality, designed to change what they see and therefore buy.				

Name	Level 1	Level 2	Level 3	Source
Fake Button	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Long et al., 2023
Definition: The user is presented with an element that appears to be a useful button, but is actually a disguised element for causing an undesirable outcome.				
False Hierarchy	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Gray et al., 2018
Definition: The user is presented with one or more options where they are given higher visual or interactive precedence than others.				
Fuzzy Targeting	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Wu et al., 2022
Definition: The user is shown products in a way that it seems to apply to any and all target populations.				
Inconsistent Content	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Long et al., 2023
Definition: The user is presented with an element that entices with an offer or benefit, but upon interacting the element fails to fulfill expectations.				
Induced Icon	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Long et al., 2023
Definition: The user is presented with icons that induce following a particular path and interact with other elements that may lead to undesirable outcomes.				
Interface Interference	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Gray et al., 2018
Definition: The user is presented with an interface that privileges specific actions over others.				
Low Contrast	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Conti and Sobiesk, 2010
Definition: The user is offered options where desirable options (undesirable to the service) are presented with low contrast.				
Mask User Warning Messages	Information Asymmetry	Active Misleading Actions	Misleading Presentation	Conti and Sobiesk, 2010
Definition: The user is prevented from viewing browser status and warning messages by the designer.				

<table>
<tr><th>Name</th><th>Level 1</th><th>Level 2</th><th>Level 3</th><th>Source</th></tr>
<tr><td>Misleading Experience Marketing</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Presentation</td><td>Mhaidli and Schaub, 2021</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is presented with a digital representation of a product through extended reality (XR) that purports to represent the real version, but may be manipulated to be better than reality.</td></tr>
<tr><td>Overlapped Placement</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Presentation</td><td>Long et al., 2023</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is shown undesirable elements that obscure or interfere with desired elements.</td></tr>
<tr><td>Trick Question</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Presentation</td><td>Mathur et al., 2019</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is misled into taking an action, due to the presentation of confusing or misleading language.</td></tr>
<tr><td>Twist</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Presentation</td><td>Kitkowska, 2023</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is presented with colours and symbols that misguide them.</td></tr>
<tr><td>Undeclared Acts</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Presentation</td><td>Long et al., 2023</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is presented with an element that appears to be a useful button, but is actually a disguised element for causing an undesirable outcome</td></tr>
<tr><td>Visual Interference</td><td>Information Asymmetry</td><td>Active Misleading Actions</td><td>Misleading Presentation</td><td>Mathur et al., 2019</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user expects to see information presented in a clear and predictable way on the page, but it is hidden, obscured or disguised.</td></tr>
<tr><td>Ad Drop-down Delay</td><td>Information Asymmetry</td><td>Passive Misleading Omissions</td><td>Delaying Provision</td><td>Lacey et al., 2023</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is presented with a delayed drop-down advertisement, leading them to accidentally click it instead of their desired action.</td></tr>
<tr><td>Autoplay</td><td>Information Asymmetry</td><td>Passive Misleading Omissions</td><td>Delaying Provision</td><td>Roffarello and Russis, 2022</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is shown content that automatically plays without the user's interaction.</td></tr>
</table>

<table>
<tr><th>Name</th><th>Level 1</th><th>Level 2</th><th>Level 3</th><th>Source</th></tr>
<tr><td>Delay User's Work Effort</td><td>Information Asymmetry</td><td>Passive Misleading Omissions</td><td>Delaying Provision</td><td>Conti and Sobiesk, 2010</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is forced to view and wait for an advertisement.</td></tr>
<tr><td>Hidden Costs</td><td>Information Asymmetry</td><td>Passive Misleading Omissions</td><td>Delaying Provision</td><td>Brignull, 2010</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is enticed with a low advertised price. After investing time and effort, they discover unexpected fees and charges when they reach the checkout.</td></tr>
<tr><td>Infinite Scrolling</td><td>Information Asymmetry</td><td>Passive Misleading Omissions</td><td>Delaying Provision</td><td>Roffarello and Russis, 2022</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user can scroll the service infinitely, with new content constantly loading.</td></tr>
<tr><td>Interactive Hooks</td><td>Information Asymmetry</td><td>Passive Misleading Omissions</td><td>Delaying Provision</td><td>Mildner et al., 2023</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user is induced to remain on the service by delayed gratification tactics.</td></tr>
<tr><td>Pull-to-refresh</td><td>Information Asymmetry</td><td>Passive Misleading Omissions</td><td>Delaying Provision</td><td>Roffarello and Russis, 2022</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user can "pull" the interface to load more content.</td></tr>
<tr><td>Centralize</td><td>Information Asymmetry</td><td>Passive Misleading Omissions</td><td>Hiding Information</td><td>Bösch et al., 2016</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user's data is collected in a single centralised location to preserves links between different users.</td></tr>
<tr><td>Comparison Obfuscation</td><td>Information Asymmetry</td><td>Passive Misleading Omissions</td><td>Hiding Information</td><td>National Commission on Informatics and Liberty (CNIL), 2020</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user struggles to compare products because features and prices are combined in a complex manner, or because essential information is hard to find.</td></tr>
<tr><td>Disguised Data Collection</td><td>Information Asymmetry</td><td>Passive Misleading Omissions</td><td>Hiding Information</td><td>Greenberg et al., 2014</td></tr>
<tr><td colspan="5"><b>Definition:</b> The user's data is gathered and used to build a rich user profile, without the user's consent.</td></tr>
</table>

Name	Level 1	Level 2	Level 3	Source
Hidden Information	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Gray et al., 2018
Definition: The user may have access to desirable options or content, but it is hidden.				
Immortal Accounts	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Bösch et al., 2016
Definition: The user deletes their account, but their associated data is kept.				
Intermediate Currency	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Gray et al., 2018
Definition: The user is encourage to buy virtual currency to spend on services, which hides the true cost in real money.				
Maximize	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Bösch et al., 2016
Definition: The user's data is collected, more than is needed to provide functionality.				
Preserve	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Bösch et al., 2016
Definition: The user's aggregated data can be deanonymized to recover relationships between persons.				
Price Comparison Prevention	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Brignull, 2010
Definition: The user struggles to compare products because features and prices are combined in a complex manner, or because essential information is hard to find.				
Shadow User Profiles	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Bösch et al., 2016
Definition: The user is represented in a server's database for a service they have never registered for.				
Social Brokering	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Mildner et al., 2023
Definition: The user's relationship to other parties on the service is never forgotten, despite the relationship being dissolved in reality.				

Name	Level 1	Level 2	Level 3	Source
Unintended Relationships	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Greenberg et al., 2014
Definition: The user's relationship to other parties on the service is never forgotten, despite the relationship being dissolved in reality.				
We Never Forget	Information Asymmetry	Passive Misleading Omissions	Hiding Information	Greenberg et al., 2014
Definition: The user's relationship to other parties on the service is never forgotten, despite the relationship being dissolved in reality.				
Attention Diversion	Free Choice Repression	Undesirable Imposition	Forced Acceptance	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user's attention is strategically targeted and kept by the service.				
Attention Grabber	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Greenberg et al., 2014
Definition: The user's attention is strategically targeted and kept by the service.				
Automating the User Away	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Gray et al., 2020
Definition: The user does not give consent or confirmation, but the service automatically performs tasks.				
Bad Defaults	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Bösch et al., 2016
Definition: The user unknowingly accepts defaults that share more personal information than they would otherwise intend.				
Bait and Switch	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Brignull, 2010
Definition: The user performs an action expecting a certain result, only to have it cause a different, likely undesired result.				
Bundled Consent	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Bongard-Blanchy et al., 2021
Definition: The user is automatically marked as consenting to multiple settings when consenting to only a single setting.				

Name	Level 1	Level 2	Level 3	Source
Captive Audience	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Greenberg et al., 2014
Definition: The user engages in an activity that takes time and the service takes advantage of this time to begin an unsolicited action.				
Default Sharing	Free Choice Repression	Undesirable Imposition	Forced Acceptance	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user unknowingly accepts defaults that share more personal information than they would otherwise intend.				
Disguised Ad	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Brignull, 2010
Definition: The user mistakenly believes they are clicking on an interface element or native content, but it is actually a disguised advertisement.				
Disguised Layout	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Long et al., 2023
Definition: The user is presented with advertisements that appear as normal content.				
Display Controversial Content	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Conti and Sobiesk, 2010
Definition: The user is unexpectedly presented with shocking content without their consent.				
Easy Trigger	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Long et al., 2023
Definition: The user can unintentionally trigger an action by virtue of overly sensitive interaction mechanisms.				
False Continuity	Free Choice Repression	Undesirable Imposition	Forced Acceptance	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is required to provide their email address to perform an action, which then automatically subscribes them to a newsletter.				
Forced Consent	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Bongard-Blanchy et al., 2021
Definition: The user is coerced into accepting fixed legal terms in exchange for access to the service.				

Name	Level 1	Level 2	Level 3	Source
Forced Continuity	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Brignull, 2010
Definition: The user is automatically charged for a service after it expires.				
Forced Enrolment	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Mathur et al., 2019
Definition: The user is automatically enrolled to an undesired component when accepting a desired component.				
Forced Viewing	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Conti and Sobiesk, 2010
Definition: The user is presented with news stories that are actually advertisements.				
Forced Wholesale	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Wu et al., 2022
Definition: The user is required to buy multiple units of a product as they have no choice to buy a single unit.				
Hidden Subscription	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Mathur et al., 2019
Definition: The user is charged a recurring fee under the pretence of a one-time fee or free trial.				
Hyper-sensitive Interface Elements	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Conti and Sobiesk, 2010
Definition: The user is unexpectedly shown an advertisement as a result of overly large mouse rollover activation regions.				
Illusion of Control	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Norwegian Consumer Council, 2018
Definition: The user is lulled into a false sense of security regarding their privacy and is then more likely to to disclose sensitive information.				
Impenetrable Wall	Free Choice Repression	Undesirable Imposition	Forced Acceptance	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is prevented from accessing a service unless they consent to perform an undesirable action.				

Name	Level 1	Level 2	Level 3	Source
Interrupt Acts	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Long et al., 2023
Definition: The user's flow is interrupted by pop-up advertisements.				
Milk Factor	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Greenberg et al., 2014
Definition: The user is forced to move through a specific work flow in order to access a service.				
Obscure	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Bösch et al., 2016
Definition: The user has great difficulty or even prevented from learning how their personal data is collected, stored, or processed.				
Preselection	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Greenberg et al., 2014
Definition: The user is presented preselected options that may not be in their interest to select.				
Privacy Zuckering	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Brignull, 2010
Definition: The user is tricked into sharing more information about themselves than they intend.				
Silent Or Invisible Behaviour	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Conti and Sobiesk, 2010
Definition: The user has additional software unknowingly installed by a service.				
Sneak into Basket	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Brignull, 2010
Definition: The user has items automatically added to their online shopping cart, without their knowledge.				
Spoof Content	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Conti and Sobiesk, 2010
Definition: The user is presented with new stories that are actually advertisements.				

Name	Level 1	Level 2	Level 3	Source
Video / Animation / Blinking / Motion / Audio	Free Choice Repression	Undesirable Imposition	Forced Acceptance	Conti and Sobiesk, 2010
Definition: The user's attention is attracted to advertisements by various visual and auditory distractions.				
Blaming the Individual	Free Choice Repression	Undesirable Imposition	Pressure Imposing	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is made to feel guilty about their choices.				
Confirmshaming	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Brignull, 2010
Definition: The user is emotionally manipulated into doing something that they would not otherwise have done.				
Continued Email Communication	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Kelly and Rubin, 2024
Definition: The user is sent one or more emails after disabling an account in an attempt to convince them to reactivate.				
Countdown Timers	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Mathur et al., 2019
Definition: The user is presented with a heightened sense of immediacy by a service imposing a deadline.				
Egoistic Norms	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Wu et al., 2022
Definition: The user is pressured to embrace norms promoted by a service.				
FoMO-centric Dark Design	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Westin and Chiasson, 2021
Definition: The user is emotionally manipulated to perform specific actions by a service leveraging its data collection and deep learning capabilities.				
Hyperpersonalization	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Mhaidli and Schaub, 2021
Definition: The user is emotionally manipulated to perform specific actions by a service leveraging its data collection and deep learning capabilities.				

Name	Level 1	Level 2	Level 3	Source
Improving the Experience	Free Choice Repression	Undesirable Imposition	Pressure Imposing	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is encouraged to share more data by the service giving an argument that it will improve the experience.				
Inducements to Reconsider	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Kelly and Rubin, 2024
Definition: The user is pressured to remain using a service through language, visuals, or incentives.				
Inducing Artificial Emotions	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Mhaidli and Schaub, 2021
Definition: The user is presented an emotive experience via extended reality (XR) that, if positive, may bias toward a positive evaluation of the service.				
Last Minute Consent	Free Choice Repression	Undesirable Imposition	Pressure Imposing	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is pressure into providing consent when the service knows the user is in a weak position due to hurry and impatience.				
Last Minute Solutions	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Kelly and Rubin, 2024
Definition: The user, when attempting to disable their account, is presented with options that the service has predicted will counteract the user's reasons.				
Making Personal Information Public	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Greenberg et al., 2014
Definition: The user's personal information is made publicly visible when the user enters a particular area of the service.				
Misleading Text	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Long et al., 2023
Definition: The user is emotionally manipulated into doing something that they would not otherwise have done.				
Nagging	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Gray et al., 2018
Definition: The user tries to do something, but they are persistently interrupted by requests to do something else that may not be in their best interests.				

Name	Level 1	Level 2	Level 3	Source
Playacting	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Wu et al., 2022
Definition: The user is pressured to purchase via a fabricated emotional story or sympathy.				
Pressured Selling	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Mathur et al., 2019
Definition: The user is steered toward options that are more desirable to the service by high-pressure tactics such as upselling and cross-selling.				
Providing Option	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Kelly and Rubin, 2024
Definition: The user is given an option to reactivate their account, either temporarily or indefinitely.				
Publish	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Bösch et al., 2016
Definition: The user's personal information is made publicly visible when the user enters a particular area of the service.				
Recommendations	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Roffarello and Russis, 2022
Definition: The user is algorithmically encouraged to consume recommended content, effectively trapping them into an endless supply.				
Repetitive Incentive	Free Choice Repression	Undesirable Imposition	Pressure Imposing	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is repeatedly offered incentives by the service to encourage them to share more data.				
Retaining Customers	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Wu et al., 2022
Definition: The user is incentivised to remain on the service longer as the designer is aware that this makes the user more likely to make a purchase.				
Rewards and Punishment	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Norwegian Consumer Council, 2018
Definition: The user is enticed to make certain choices over others by being rewarded for making a designer-aligned choice and punished for others.				

Name	Level 1	Level 2	Level 3	Source
Safety Blackmail	Free Choice Repression	Undesirable Imposition	Pressure Imposing	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is pressured into consenting to unnecessary sensitive data collection under the false pretence of extra security.				
Social Investment	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Roffarello and Russis, 2022
Definition: The user is captured by social metrics such as reactions, comments, followers, to “bind” them to the service.				
Social Pyramid	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Gray et al., 2018
Definition: The user is incentivised to recruit other users to the service.				
Targeting Vulnerable Consumers	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Mhaidli and Schaub, 2021
Definition: The user is personally targeted by an algorithm with personal knowledge of their vulnerabilities.				
Threatening Messages	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Conti and Sobiesk, 2010
Definition: The user is prompted to perform an action as a result of receiving a threatening message.				
Toying With Emotion	Free Choice Repression	Undesirable Imposition	Pressure Imposing	Gray et al., 2018
Definition: The user is emotionally manipulated by the service’s use of design feature to take particular actions.				
Bury in Navigation Hierarchy	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Conti and Sobiesk, 2010
Definition: The user is hindered from finding and using desired actions by hiding them in an unnecessarily complicated navigation hierarchy.				
Complete Obstruction	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Kelly and Rubin, 2024
Definition: The user is completely prevented from completing desired actions, such as deleting an account.				

Name	Level 1	Level 2	Level 3	Source
Contact Zuckering	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Lacey et al., 2023
Definition: The user is obstructed from finding the organisation’s telephone number.				
Controlling	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Gray et al., 2020
Definition: The user is restricting from following their own task flow and is instead explicitly directed to follow the designer’s.				
Decision Uncertainty	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Mildner et al., 2023
Definition: The user is made to feel unsure about what is expected of them or what options are available.				
Deny	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Bösch et al., 2016
Definition: The user is denied control over their data.				
Ease	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Norwegian Consumer Council, 2018
Definition: The user is lead in a certain direction, usually aligned with the designer’s intentions, and alternatives are a long and arduous process.				
Entrapping	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Gray et al., 2020
Definition: The user is mislead by the design and falls into a trap that cannot be avoided or corrected.				
Forced Email Confirmation	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Kelly and Rubin, 2024
Definition: The user is required to confirm their choice to disable their account by responding to an email.				
Forced Explanation	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Kelly and Rubin, 2024
Definition: The user is required to select or write a reason for performing a desired action before the service will permit them.				

Name	Level 1	Level 2	Level 3	Source
Gamification	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Gray et al., 2018
Definition: The user is only able to access certain aspects of a service through “grinding” or else purchase upgrades.				
Hard to Cancel	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Mathur et al., 2019
Definition: The user is given very easy options for signing up to a service, but is obstructed from cancelling.				
Hide Desired Interface Elements	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Conti and Sobiesk, 2010
Definition: The user's desired action is placed in an obscure location to maximise advertisement view time.				
Hinder Confidential Settings	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is able to consent with a simple action, but the process of data protection is long and complicated.				
Labyrinthine Navigation	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Mildner et al., 2023
Definition: The user is presented with nested interfaces that are easy to get lost in, disabling users from choosing preferred settings.				
Make Uninstalling Difficult	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Conti and Sobiesk, 2010
Definition: The user is prevented from performing a desired action, such as uninstalling.				
Missing Exit	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Long et al., 2023
Definition: The user is prevented from exiting an interface through easy means, leading them to more easily select an option preferred by the designer.				
Obfuscating Settings	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	National Commission on Informatics and Liberty (CNIL), 2020
Definition: The user is forced to go through a deliberately long and tedious process to achieve the setting they desire.				

Name	Level 1	Level 2	Level 3	Source
Obstruction	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Gray et al., 2018
Definition: The user is impeded from their task flow by a design that has the intent to dissuade that task flow.				
Omit Necessary Controls	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Conti and Sobiesk, 2010
Definition: The user is prevented from performing desired actions by the service lacking the relevant control.				
Requiring Request	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Kelly and Rubin, 2024
Definition: The user must submit a request for account disabling, which must then be approved.				
Restricted Options	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Ahuja and Kumar, 2022
Definition: The user is forced by the design functionality or choice architecture to choose from a set of choices that bar the most relevant, optimal, or desirable ones.				
Roach Motel	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Brignull, 2010
Definition: The user finds it easy to sign up or subscribe, but when they want to cancel they find it very hard.				
Temporary Obstruction	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Kelly and Rubin, 2024
Definition: The user is forced to take actions that are not inherently necessary to their desired action, which increases their workload.				
Typing Errors	Free Choice Repression	Undesirable Restriction	Restricting Specific Actions	Conti and Sobiesk, 2010
Definition: The user is presented with an advertisement instead of assistance when they make a mistake, such as mistyping a URL.				
Forced Action	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Brignull, 2010
Definition: The user wants to do something, but they are required to do something else undesirable in return.				

Name	Level 1	Level 2	Level 3	Source
Forced Endorsement	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Wu et al., 2022
Definition: The user wants to obtain a desirable reward or perk from the service, but must first perform an action desirable to the service.				
Forced Registration	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Bösch et al., 2016
Definition: The user is required to make an account and give personal information in order to access the service.				
Mandatory Form Field Entries	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Conti and Sobiesk, 2010
Definition: The user is required to enter contact information before they are allowed to accomplish the task.				
Nickling-and-diming	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Gray et al., 2020
Definition: The user is prevented from interacting with a service by an initially disguised requirement for payment.				
Pressure to Receive Marketing	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Kitkowska, 2023
Definition: The user must opt into receiving marketing in order for the service to allow them to register.				
Redirective Conditions	Free Choice Repression	Undesirable Restriction	Restricting Specific Users	Mildner et al., 2023
Definition: The user is required to overcome unnecessary obstacles before being able to achieve their goals.				

Appendix B

The Python code used to query OpenAI's GPT-4 GAI LLM.

```
import csv
import pandas as pd
from openai import OpenAI

'''
The purpose of this code is to evaluate the efficacy of CDR standards to
deceptive patterns identified in the IVE Deceptive Patterns Typology. The
code incorporates gpt4-API and selected information of CDR standards and
the
IVE Deceptive Patterns Typology. The code provides the selected
information to gpt4 and retrieves the response which indicates the
efficacy of the CDR standard to the deceptive pattern.

Input: Two files for CDR standards and the IVE Deceptive Patterns
Typology.
Output: A result file.
'''

client = OpenAI(api_key='...')
file_path = './'

# Input files
standards = pd.read_csv(file_path + 'standards.csv')
patterns = pd.read_csv(file_path + 'patterns.csv')

with open('result.csv', 'w', newline='', encoding='utf-8') as file:
     writer = csv.writer(file)
     writer.writerow(['DP-ID', Deceptive pattern name', 'Efficacy of CX
rule', 'CX-ID', 'Focus area', 'Type', 'Participant',
                      'Requirement', 'CX statement'])

     # For each deceptive pattern
     for index, pattern in patterns.iterrows():
     DP_ID = pattern['DP-ID']
     pattern_definition = pattern['Definition']
     deceptive_pattern_name = pattern['Name']

     # For each CDR standard
     for idx, rule in cx_rules_filtered.iterrows():
          CX_ID = rule['CX-ID']
          focus_area = rule['Focus area']
          rule_type = rule['Type']
          participant = rule['Participant']
          requirement = rule['Requirement']
```

```
           statement = rule['Statement']

           # Invoke gpt4-API
           message = [{"role": "system",
                      "content":
                           f"You are a helpful assistant "
                           f"Given the consumer experience rule:
{statement} in the focus ares: {focus_area} with "
                           f"requirement: {requirement} that
participants: {participant} follow,"
                           f"can this consumer rule address the dark
pattern named: {deceptive_pattern_name} which has "
                           f"characteristic of {pattern_definition}"
                           f"and reduce the risk associated with the
dark pattern: {deceptive_pattern_name}?"
                           f"Respond only with Yes or No."
                           f"If the consumer experience rule cannot
address any dark pattern, you must respond "
                           f"with No."
                      }]

           # Obtain response from gpt4-API
           response = client.chat.completions.create(
                model="gpt-4",
                messages=message,
                temperature=1.0,
                max_tokens=50
           )

           # Record response in result file
           if 'Yes' in response.choices[0].message.content:
                efficacy = 'yes'
           else:
                efficacy = 'no'
           writer.writerow(
                [DP_ID, deceptive_pattern_name, efficacy, CX_ID, focus_
area, rule_type, participant, requirement, statement])
```