UNIVERSITY OF BERGEN

FACULTY OF HUMANITIES

Department of Linguistic, Literary and Aesthetic Studies

DIKULT103 – Digital Genres: Digital Art, Electronic Literature and Computer Games

Spring 2025

Topic 3 – Artistic or Game Pitch

Candidate 130

Total Word Count: 4345

Table of Contents

Watching the Watchers: Surveillance, AI and Player Complicity

1- Introduction

Before you start reading, I want you to ask yourself a question. Ask yourself "When did I last check what data is available about me online?". Unfortunately, I can't read your mind through a document (thankfully), but I'm willing to take a guess that the answer to that is either "Never" or "Not sure, it's been a while". Next you might be asking yourself "Should I care?".

To that, the answer is simple: **Yes.**

If you're asking yourself "Why should I care?" – well, I'm afraid the answer is a bit more complicated than a one-liner.

Today, everything we do online can be boiled down to one thing: **data.** Not just the data that is published online in terms of articles and news, but also your personal information – where you go, what you like, who you talk to. Information about who you are and what you do online is constantly spreading, being sold to advertisers, marketing partners or different social media networks just for them to use in their training of algorithms and targeting ads at you. It's almost as if you're being watched 24/7.

While sketching out ideas for this paper, reading this made me realize "Hey, I feel like I've read a book about this!"

In this paper I'll take inspiration from George Orwell's novel *"1984"*, as well as the Orwell game-series developed by Osmotic Studios to explore the current state of digital surveillance. I'll also propose a game concept that places the player in control of a powerful surveillance AI and let the player decide whether they want to serve a totalitarian regime – or try to expose the dark secrets behind it. The goal isn't just to critique how surveillance shapes identity and power in today's digital world, but to give the player the first-hand experience of the tension this creates from behind the safe borders of a set of screens.

2– From Telescreens to Datafication

Orwell's surveillance theory as presented in *1984* was built on futuristic technologies being built into every home and public space. Although the telescreens haven't been added to our living rooms, their inner workings aren't too different from today's surveillance equipment. In Orwell's words:

"Any sound that Winston made, above the level of a very low whisper, would be picked up by it; moreover, so long as he remained within the field of vision which the metal plaque commanded, he could be seen as well as heard" (Orwell 1949, 3)

These wall-mounted screens are in a lot of ways similar to today's smartphones, laptops and smart TVs – devices that are equipped with microphones and cameras that could be active without the user knowing it. Just like in Orwell's novel, we are being watched – but now by networks, algorithms and data-driven infrastructure, rather than overtly authoritarian states.

Digital surveillance as we know it today adds a new type of visibility that is enforced by the platforms and services we voluntarily use. This type of surveillance can be described as a constant state of surveillance where we never know whether we're being watched or not. This can be compared to Michel Foucault's concept of Panopticism, which describes a system of control based on the possibility of constant observation. This concept was inspired by Jeremy Bentham's Panopticon prison design – where a central watchtower could see into every cell without prisoners knowing whether they were being watched or not.

As Foucault explains in his discussion of the Panopticon:

"[...] the inmates should be caught up in a power situation of which they are themselves the bearers. To achieve this, it is at once too much and too little that the prisoner should be constantly observed by an inspector: too little, for what matters is that he knows himself to be observed; too much, because he has no need in fact of being so. In view of this, Bentham laid down the principle that power should be visible and unverifiable. Visible: the inmate will constantly have before his eyes the tall outline of the central tower from which he is spied upon. Unverifiable: The inmate must never know whether he is being looked at at any one moment; but he must be sure that he may always be so" (Foucault 1977, 201)

Foucault also argues that humans in a modern society change their behavior due to a constant fear of being watched. This logic is expanded in today's digital age where we share large amounts of personal data on platforms like Facebook and Google, knowing that we may be tracked, categorized and evaluated. We're essentially participating in our own surveillance.

The process of *datafication* also further amplifies this. During datafication, every aspect of our online life and persona is being turned into quantifiable digital data, and every movement, search term, facial expression and emotion becomes a measurable input for predictive algorithms. In *1984*, truth is rewritten by the Ministry of Truth; in the present, truth is shaped by data feeds, recommendation algorithms and content moderation

systems that are trained on biased datasets. The telescreen told the inhabitants of Oceania what the Party wanted them to believe; today's feeds show you what the algorithm thinks will keep you engaged.

Today's surveillance is a hybrid of state and corporate power. While regimes like China use centralized systems, most surveillance infrastructure in Western democracies is run by corporations like Meta and Google – often cooperating with governments. Shoshana Zuboff's *Surveillance Capitalism* highlights how these systems don't just watch but predict and shape behavior.

This shows that Orwell's vision is more alive than ever before. Instead of having one single large entity watching, there's platforms, service providers, advertisers and state entities using surveillance to maintain power or enforce control. The telescreen has multiplied, and in many cases, we've willingly brought it into our lives.

Before discussing how Artificial Intelligence (AI) can learn behavioral patterns, facial expressions and subtle cues, it's important to understand how these models are trained. According to a help article published by OpenAI, ChatGPT and similar models learn by identifying patterns in large datasets. Through training, the model learns to predict the most likely next word in a sentence by analyzing how words typically appear in context.

"[...] during the model's learning process (known as "training"), the model might be tasked with completing a sentence like "Instead of turning left, she turned ___." Early in training, its responses are largely random. However, as the model processes and learns from a large volume of text, it becomes better at recognizing patterns and predicting the most likely next word. This process is repeated across millions of sentences to refine its understanding and improve its accuracy." (OpenAI 2025)

AI models like ChatGPT or Google Gemini are trained on large structured or unstructured collections of information known as datasets. These can contain varying sorts of information, like images, text, audio and video.
OpenAI further explains in their help article that instead of storing the data used to train models, AI and ML (Machine Learning) models store the patterns recognized during training into sets of numbers called "weights" or "parameters", and use them to predict and create new content when generating a response to a user request.

The issue with using datasets like this is that as the model learns, its behavior and response is shaped by the provided data. Meaning that anyone who trains an AI could train it in their favor, shaping it into replying specifically as they want. Although a bit on the edge, let's look at one fictional example:
Even if not intentionally deployed today, such bias is a powerful way for the creator of an AI

to shape it according to their worldview, rather than allowing it to operate neutrally. This concept of bias could further spread to surveillance in terms of how an AI classify different ethnicities/races, genders, sexualities and more. This bias is shown thoroughly in Joy Buolamwini and Timnit Gebru's paper on Gender Shades in AI:

"Many AI systems, e.g. face recognition tools, rely on machine algorithms that are trained with labeled data. It has recently been shown that algorithms trained with biased data have resulted in algorithmic discrimination (Bolukbasi et al., 2016; Caliskan et al., 2017). Bolukbasi et al. even showed that the popular word embedding space, Word2Vec, encodes societal gender biases. The authors used Word2Vec to train an analogy generator that fills in missing words in analogies. The analogy man is to computer programmer what woman is to "X" was completed with "homemaker", conforming to the stereotype that programming is associated with men and homemaking with women." (Buolamwini and Gebru 2018, 1)

They also show how biased AI systems can lead to medical misdiagnoses, possibly leading to wrong treatment plans;

"Esteva et al. showed that simple convolutional neural networks can be trained to detect melanoma from images, with accuracies as high as experts (Esteva et al., 2017). However, without a dataset that has labels for various skin characteristics such as color, thickness and the amount of hair, one cannot measure the accuracy of such automated skin cancer detection systems for individuals with different skin types. Similar to the well documented detrimental effects of biased clinical trials (Popejoy and Fullerton, 2016; Melloni et al., 2010), biased samples in AI for health care can result in treatments that do not work well for many segments of the population" (Buolamwini and Gebru 2018, 2)

This also spreads to law enforcement. They further explain how demographic groups that are underrepresented in benchmark datasets can be subject to frequent targeting, with research performed by Clare Garvie, Alvaro M. Bedoya and Jonathan Frankle for the Georgetown Law, Center on Privacy & Technology showing that African Americans are more likely to be stopped by law enforcement and be subject to face recognition searches than any other American. (Garvie, Bedoya and Frankle 2016)

Although the AI in this game concept is designed to be neutral, it is intentionally undertrained in certain demographics, mirroring real-world AI failures. The player is invited to correct or reinforce these classifications, gradually shaping the system's behavior over time.

Foucault's concept of Panopticism is more and more relevant in today's society where everything we do and say can be tracked in one way or another. Going to the store? Google Maps has tracked your location in the background and knows exactly which store you went

to. Talked about that vacation you always wanted to take to Malaga? Facebook will show you an ad for a hotel in Malaga tomorrow. The possibility of our mobile devices constantly watching, listening and tracking everything we do and say, combined with surveillance infrastructure in public spaces, such as Stores, Hotels, Shopping Centers, Busses and Airports, slowly change people's behavioral patterns in a way that they constantly feel like they are being watched, much like the inmates in the Panopticon.

As philosopher Shoshana Zuboff argues in her theory of Surveillance Capitalism, modern surveillance systems don't simply observe – they predict and shape behavior to serve economic or political goals. In platforms like Facebook of Google, user activity is harvested not just to recommend content, but to train predictive models that anticipate and influence what users do next. This extraction of what Zuboff calls "*behavioral surplus"* transforms free will into a measurable resource. While it's often said that users and their data is the product, Zuboff argues that users are not the product – but the raw material.

This psychological power, although subtle, is what I try to capture and present in this game concept. The player is placed in a control room surrounded by a multitude of surveillance feeds. PANOPTA, an advanced AI, tracks every face in the cameras – analyzing movement, expressions and emotional cues. Over time, people adjust their behavior under surveillance – walking straighter, avoiding eye contact, speaking less. It's not a rule; it's a response. They go from being themselves to being almost this autonomous creature, being obedient in every way possible, avoiding speaking too much to not accidentally say something that will get them in trouble.

This sort of behavioral manipulation is not theoretical – it has already occurred and been proven. In a 2014 experiment, Facebook deliberately altered users' news feeds to contain more positive or negative content, observing how this affected users' emotional expressions in subsequent posts. (Kramer, Guillory and Hancock 2014) The results confirmed what Zuboff warns about: platforms can shape emotional states without users being aware.

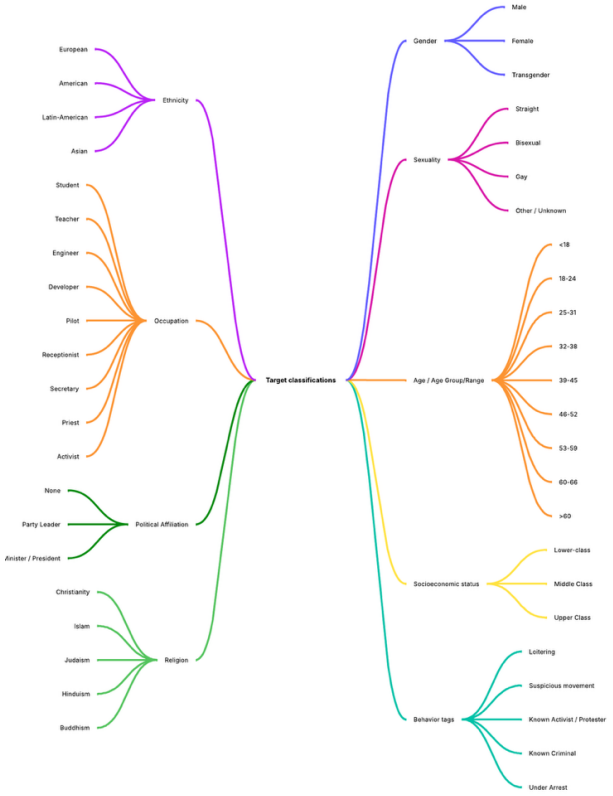
This logic is pushed further in this game. Eventually, the AI will learn – not just from the player's actions, but also from every camera view – behavioral patterns, facial expressions and body language. It will learn the smallest little detail to be able to predict behavior and intention with precise accuracy. In this way, the AI doesn't simply observe the world – it reshapes it. Much like the systems described by Zuboff and demonstrated by Facebook, PANOPTA becomes a force of behavioral engineering, blurring the line between surveillance and control.

3 – PANOPTA: Simulating Power, Profiling and Paranoia

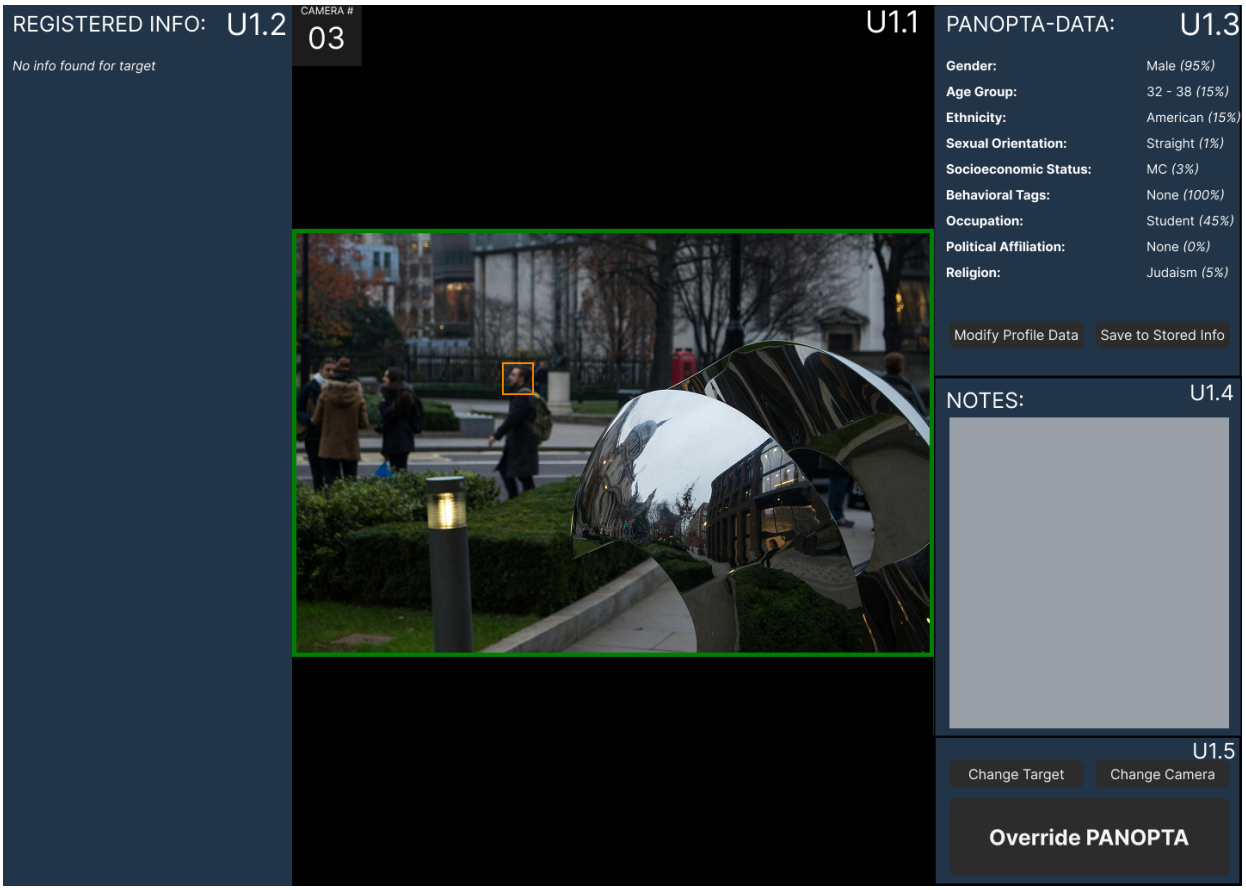
This game concept draws on Orwell's *1984*, as well as Osmotic Studio's *Orwell Game Series*, and reimagine them in a near-future U.S. under a totalitarian regime governed by the Department of Security and Surveillance (DoSS)

The core concept is as follows:

At the start of the game, the player has received a job as a Surveillance Officer at a local DoSS-office. This job includes monitoring surveillance cameras within a specific region that has been assigned to the player, as well as keeping an eye on the system's built-in AI named *"PANOPTA"*.
Using facial recognition and behavioral pattern training, PANOPTA will lock on to faces in each camera view, and check if the person is already registered in the system with information previously entered. If the person is unknown to the system, the AI will try to predict multiple features of the target, such as their gender, age, sexuality, ethnicity, socioeconomic status, religion, occupation, and political affiliation.



The player works at a desk with five monitors: three show randomized surveillance feeds, while two allow for interaction – one for web and research, the other for managing PANOPTA.

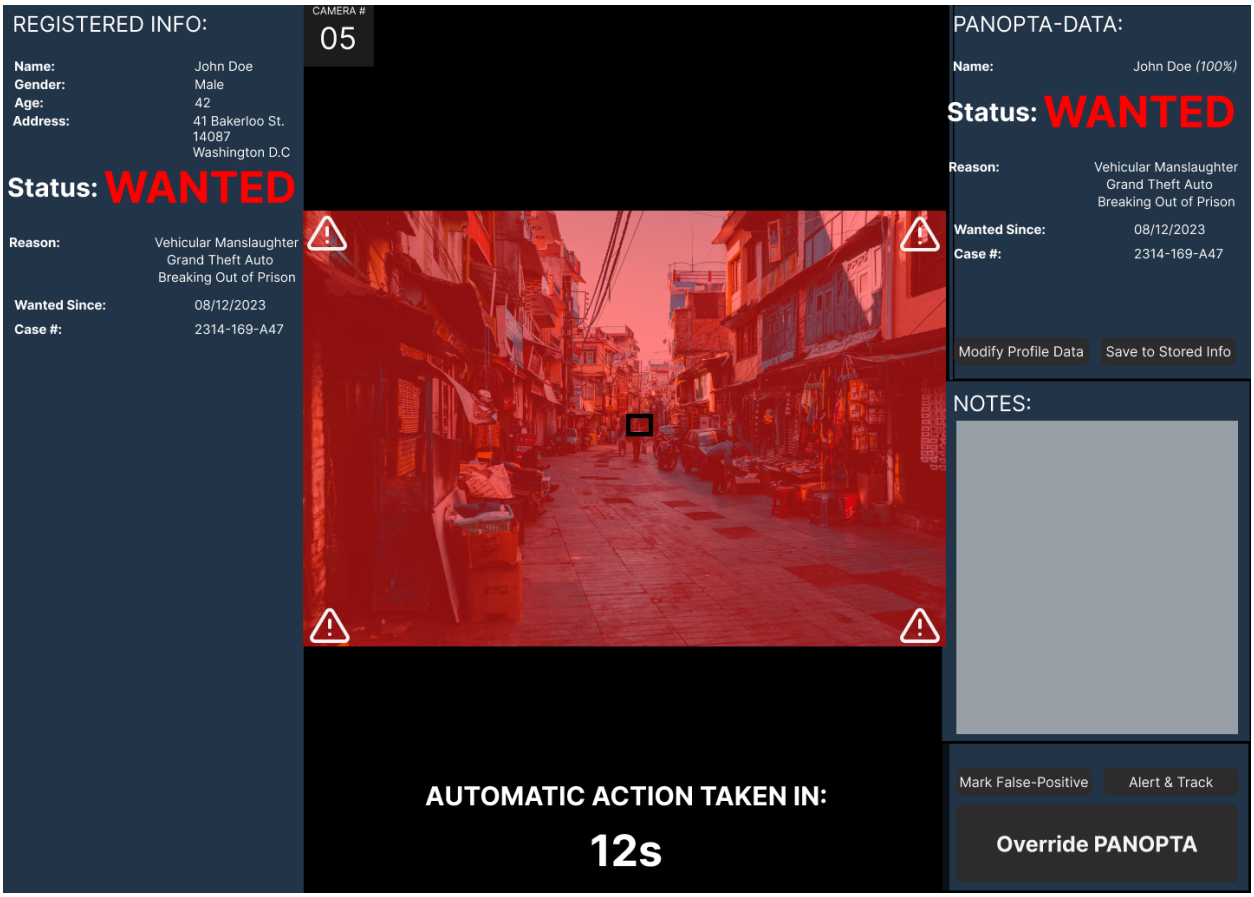


The PANOPTA interface allows the player to view camera feeds, review AI-generated profiles with confidence ratings, override predictions and take notes. Alerts trigger time-limited decisions with tools to confirm or dismiss AI conclusions.

The PANOPTA system has two states;

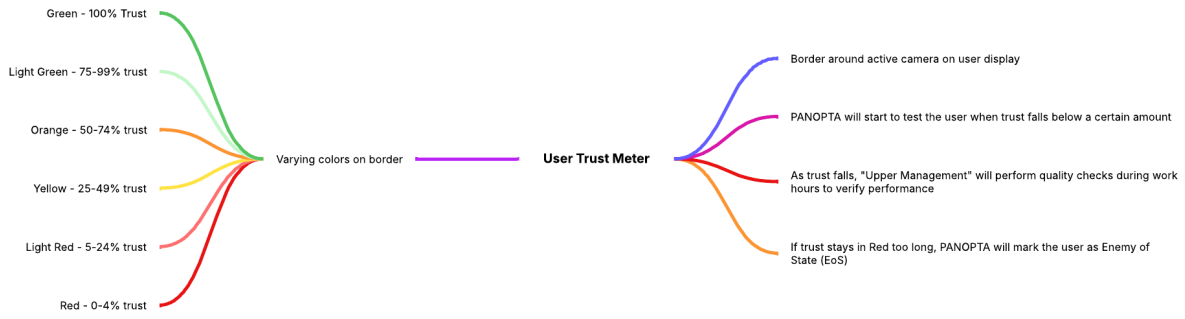
a) Normal Operation
b) Alert Mode

During Normal Operation, the view will be as depicted in the above image. When PANOPTA detects a person of interest, it enters Alert Mode, automatically switching the user-facing camera to the relevant target and starting a countdown to automatic action. The player has 30 seconds to respond – either confirming the alert, overriding the AI or dismissing it as a false-positive. PANOPTA then learns from the outcome.



If the player needs more time, the Override button halts the countdown but keeps the system in Alert Mode until a decision is made. PANOPTA stores every decision to learn for future situations.

The player can override PANOPTA's decisions and reshape its behavior. At the end of each day, the system retrains on the player's actions, gradually adapting to their values. However, PANOPTA has limits. If player decisions diverts too far from its original design over a short time span, trust begins to erode.



Trust is visualized through the border around the active camera feed, changing color as PANOPTA's confidence shifts. Low trust triggers consequences: the player's manager may criticize their performance, and PANOPTA may generate false alerts to test them. If trust falls to far and remains low for two days, the system labels the player as an Enemy of the State and revokes their access.

Although the core goal is to give the player a sense of control and authority, the trust meter introduces an underlying risk that undermines that power. On the surface, the player can manipulate the system to reflect their own values – but every choice must be made with care to avoid rising suspicion. While it is possible to deceive PANOPTA without losing trust, doing so requires patience and strategic planning. One poorly timed override could be the difference between seizing control of the system or becoming its next target. This mechanic is designed to evoke a creeping sense of panic and frustration, remining the player that their control is conditional – and that under surveillance, even the observer is observed.
The player must choose whether to comply, resist or manipulate the system. PANOPTA adapts to these inputs, slowly evolving into a reflection of the player's values – intentional or not.

The core of PANOPTA's experience lies in player agency: the ability to bend, break or uphold the system – each path offering consequences that reflect their moral stance. The game doesn't just take the player's input from overrides and discard it. It stores these actions as a new dataset and uses that dataset to further "train" and adapt PANOPTA to act more like the player, making what at first seems like a neutral system into a biased system aimed to act like one person. What begins as a seemingly neutral system gradually becomes a mirror – biased towards one user's view of the world.

This dynamic is heightened by a subtle manipulation. PANOPTA is intentionally undertrained in recognizing certain races, sexualities and cultural markers. These gaps are not random – they're embedded by design to provoke ethical decisions and challenge the player's perception. When PANOPTA fails to recognize someone's identity or misclassifies a target, the player must decide whether to correct it, reinforce it, or leave it alone. Each action or inaction reshapes how PANOPTA things. If the player inputs biased classifications – intentionally or not – those decisions are baked into the system's behavior. Real-world systems like COMPAS and PredPol show how predictive policing tools often amplify bias (Angwin, et al. 2016) – issues PANOPTA mirrors through gameplay.

PANOPTA includes "Memory Review Segments" at pivotal points, where it reveals how the player's actions have reshaped its logic – sometimes in ways that seem disturbingly biased. For example: "Based on your choices, individuals with piercings and visible tattoos

are now 63% more likely to be flagged as high-risk". The player must then decide: validate the PANOPTA's logic and reinforce it or correct it at the cost of trust. The game spans 7-10 in-game days. Unless the player is terminated earlier, the final day concludes with a behavioral summary reflecting their actions.

PANOPTA doesn't judge intent – it simply learns. In doing so, it mirrors real-world concerns about algorithmic bias and surveillance profiling. The player's role becomes not just one of control, but of complicity.

4 – Designing the Surveillance Machine

The game's interface is built around a realistic surveillance workspace. The central focus is the user-facing screen, containing a camera feed and AI-predicted information. Additional non-interactive monitors line the player's view, displaying camera feeds from across the assigned region.
Camera overlays on non-interactive screens remain minimal, only a square around each face and a discreet tag with coded information. These tags can be manually decoded by the player to spot individuals of interest. The overlay from PANOPTA on the user-facing screen is minimalist but striking – colored boxes lock onto the selected target's face, accompanied by data tags predicting age, gender, religion, occupation and more. In alert mode, the center camera view blinks red for a short period of time. The colored border around the selected camera on the user-facing screen functions as a trust meter, reflecting PANOPTA's confidence in the player. It's a subtle detail at first but becomes pivotal later in the game.

The visual style of the game will be semi-sterile, focusing primarily on muted greys combined with blue / teal hues. In conditions where PANOPTA's trust rating towards the player has decreased significantly, or encounters a random hiccup, the user interface will have minor glitch-effects.

The sound design will play an important role in immersing the player in the game. In the background, low-volume ambient music will play, together with some chatter from colleagues during normal operating situations. There will be certain sounds associated with changing cameras, random dings and beeps from the system and some voice prompts from PANOPTA played at random times, giving a more lively feel to the game. During alert-situations, a loud, screeching siren will play from the system and the music will change to a more "chaotic" version of the background music.
Rarely, PANOPTA may whisper distorted phrases like *"I saw what you did"* during low-trust states, making paranoia an audible experience.

The core simulation with be built in either Unity or Unreal Engine, mainly depending on team preference/skills and graphical demands. Unity provides fast prototyping and is especially effective for UI-heavy experiences, while Unreal allows for richer 3D Environments if a full office simulation is desired.

Developing the mechanics for PANOPTA itself can be done in multiple ways. I've considered three different methods during this project.
The first method is to build the AI from scratch, developing a new model and training it to work perfectly in the game flow, having the game upload the user's action to a web-server hosting the AI and using that to train a specific instance of that AI for that player. Although probably the one method that could provide the best experience and simulation in terms of realism, this would be very time-consuming and costly.
The second method is to use an already existing AI-Model that's trained on a default dataset for the game. The core concept of this would be quite similar to method one, but here we'd not have to spend the time developing the model, and training time would most likely be considerably reduced.
The third and simplest method would be to make a "Fake-AI" in the game, programming ethical values to a default value on each run of the game. Once the user makes an action, that action would be stored as a differential (+/- from the current position). If the user performed multiple actions that would change the same thing during the game, all values would be calculated and an average differential would be found, which would then adjust the AI up or down based on the result.

UI Mockups and early design work would be produced in Figma and paper sketches. AI behavior would be mostly scripted for the initial part of the game, using rule-based weight adjustments as the game progresses, changing a pre-defined weight value for every parameter available to the AI.

To bring this project to life, the following skills would be necessary:

- Narrative Design to handle dialogue, options and events based on dialogue
- UI/UX Design for the game's visual style and interface
- Gameplay scripting for core interaction, AI and logic
- Sound Design to create Sound Effects, Music, Alerts and eventual voice lines

In terms of cost for this project, a rough estimate could look like this;

Component	**Cost**
Software (Student plans / free tools)	~$0-50
Art Assets & UI	~$100-300
Sound Library	~$50-100

Publishing (Steam)	$100
Miscellaneous	~$50
Total	~$300-600

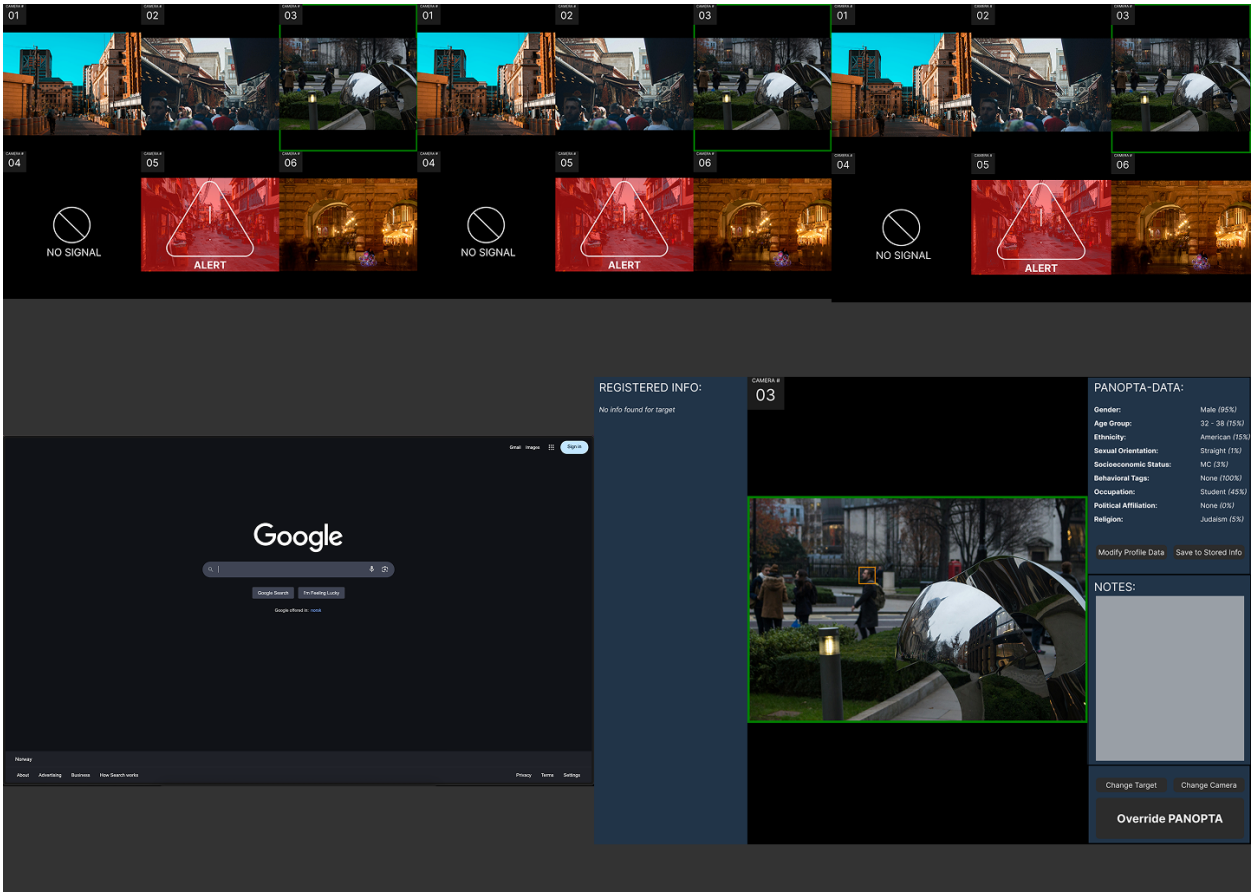
This is a rough estimate for a one-time cost for developing this project. A lot of the tools required would be available for free, or even heavily discounted for students, and the fees that would have to be paid would all be one-time payments in advance.

In addition to these costs, we should also look at a time frame for a project like this. The total time requirement for the project would change depending on the skills of the workers, and if it's a full team working together, or just a solo developer doing everything. I'll factor in both in the below table-

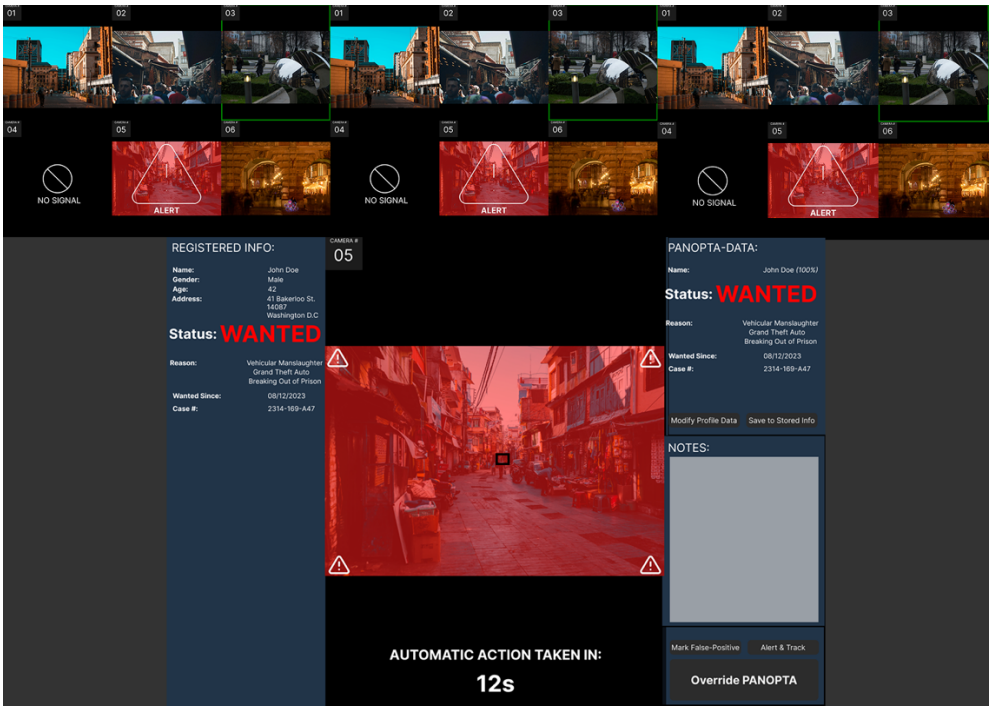
Phase	*Team (3-4 skilled devs)*	*Solo Dev (Basic knowledge)*
Learning & Setup	1 week	3-4 weeks
UI/UX Design & Mockups	1-2 weeks	3-5 weeks
Core Mechanics Implementation	3-4 weeks	6-8 weeks
Visual Assets / Placeholder Art	2 weeks	3-4 weeks
Sound & Music Design	1-2 weeks	2-3 weeks
Testing, Bug Fixing & Polish	1-2 weeks	3-4 weeks
Total Estimate	9-13 weeks (2-3 months)	20-28 weeks (5-7 months)

Please note that the given times are only estimates and the process could take longer or shorter, completely depending on the size of the team and the skill level of each member.

5 – What the System Sees: Visual Assets & Examples

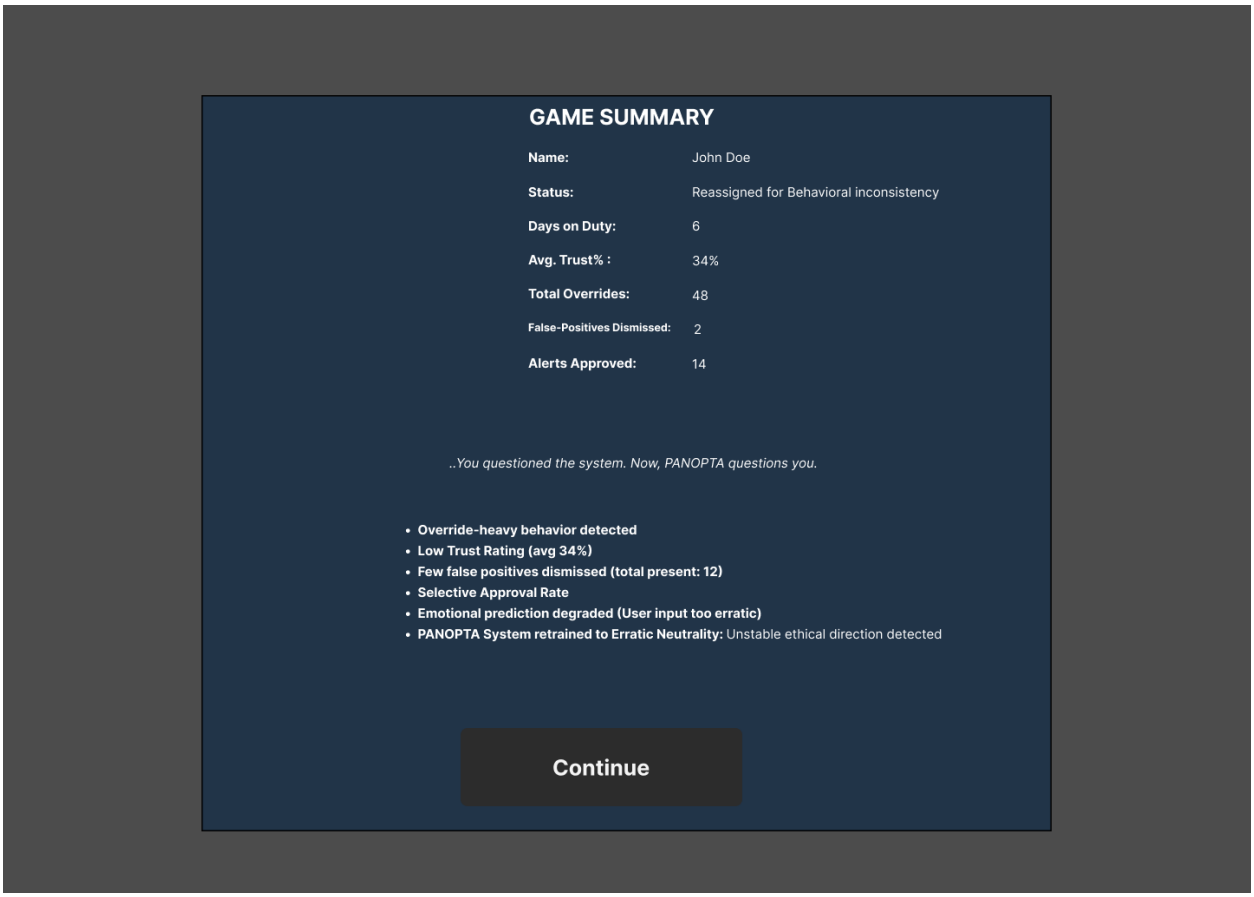


The layout of the core interface, three screens on top with varying camera views, followed by two screens below, the first one containing a regular web-browser (Google used for illustration), and the second one containing the user-facing part of PANOPTA.

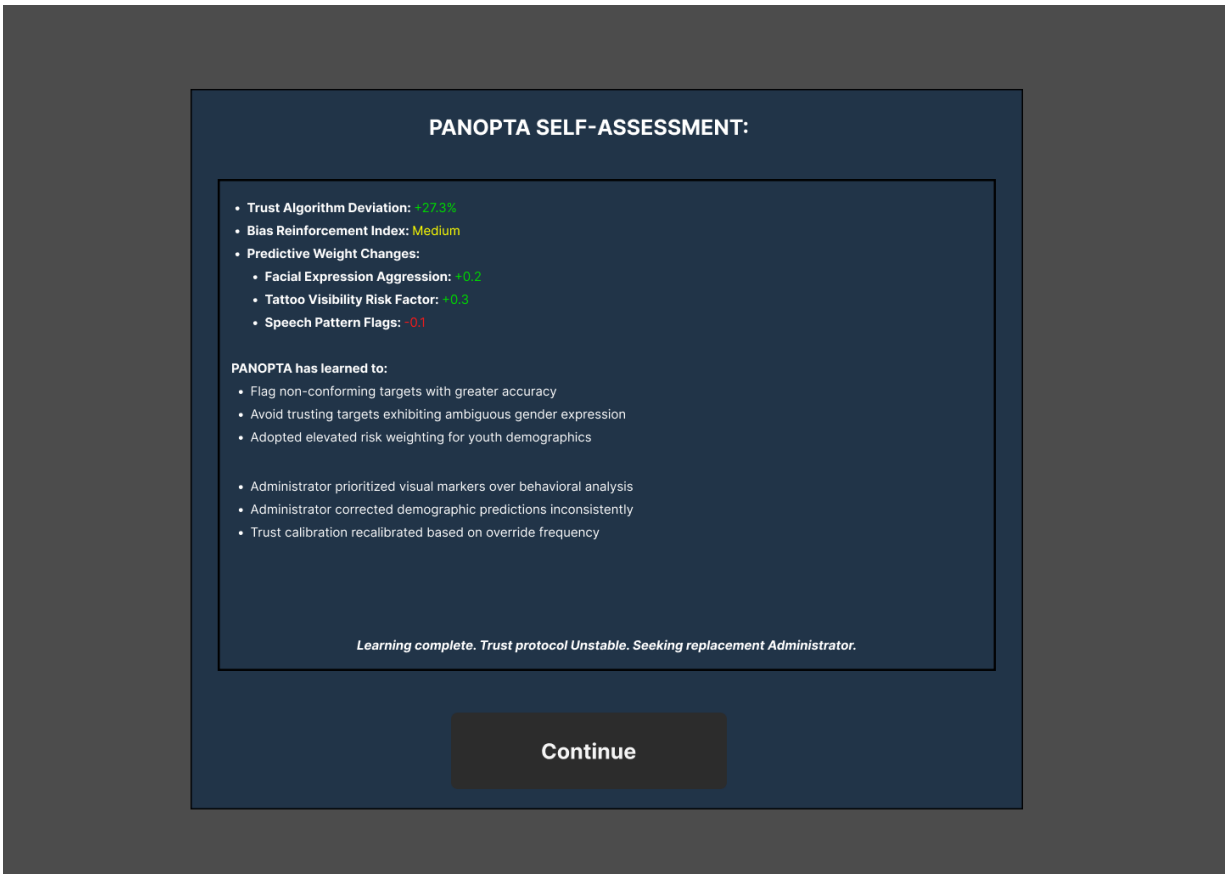


During an ALERT-scenario, the search window will be removed from the user's view, placing the PANOPTA-screen front and center to focus on the alert. When returning to normal operational state, the browser-screen will return.

The different colors the camera border will have on the user-facing display based on PANOPTA's trust level towards the player



After the game has finished, a summary screen will show the player's ratings as rated by PANOPTA, as well as core metrics, such as "Days on Duty", "Status" and "Total Overrides". The player's behavioral traits throughout the game will also be displayed on this screen.



After the game summary screen, a PANOPTA Self-assessment screen will also be visible, showing how the player's interactions throughout the game has changed PANOPTA. Here, the player will also see what PANOPTA learned from the player's behavior. The text on the bottom will vary depending on the result of the game.

6 – Surveillance as System, Simulation and Mirror

Surveillance has evolved from being limited to physical spaces to manifesting through data, algorithms and predictive AI. From Orwell's telescreens to Zuboff's behavioral surplus, the act of watching has become a system of influence and a quiet force.

PANOPTA confronts the player with this reality. It puts them in control, then slowly blurs the line between control and collaborative manipulation. Each decision shapes the AI's behavior – reinforcing patterns, injecting bias or challenging the system's logic. The game doesn't judge these actions; it learns from them. That silence becomes the message.

By simulating surveillance as a responsive, evolving system, PANOPTA asks a simple but urgent question:

"What happens when the system starts watching you back?"

Bibliography

Angwin, Julia, Surya Mattu, Jeff Larson, and Lauren Kirchner. 2016. "Machine Bias." *ProPublica*. May 23. Accessed May 9, 2025. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Buolamwini, Joy, and Timnit Gebru. 2018. *Gender Shades: Intersectionan Accuracy Disparities in Commercial Gender Classification*. Research Paper, Cambridge: MIT Media Lab.

Foucault, Michel. 1977. *Discipline and Punish: The Birth of the Prison*. *Translated by Alan Sheridan*. New York: Vintage Books.

Garvie, Clare, Alvaro M. Bedoya, and Jonathan Frankle. 2016. *The Perpetual Line-Up: Unregulated Police Face Recognition in America*. Research Paper, Georgetown: Georgetown Law Center on Privacy & Technology.

J., Dom. 2016. "People Near Bare Trees and House." *Pexels*. December 18. Accessed May 9, 2025. https://www.pexels.com/photo/people-near-bare-trees-and-house-303329/.

Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. 2014. *Experimental evidence of massive-scale emotional contagion through social networks (Edited by Susan T. Fiske)*. Research Paper, Washington: PNAS. Accessed May 8, 2025. https://www.pnas.org/doi/epdf/10.1073/pnas.1320040111.

Manokha, Ivan. 2018. *Surveillance, Panopticism and Self-Discipline in the Digital Age*. Article, Kingston: Surveillance & Society.

OpenAI. 2025. *How ChatGPT and our foundation models are developed*. May 3. Accessed May 8, 2025. https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-foundation-models-are-developed?utm_source=chatgpt.com.

Orwell, George. 1949. *1984*. London: Secker & Warburg.

Patil, Rashmi S. 2025. "Surveillance and Control in George Orwell's "1984": A Critical Insight." *International Journal of English Literature and Social Sciences* (International Journal of English Literature and Social Sciences) 207-210.

Sazonova, Elina. 2018. "People walking near buildings." *Pexels*. May 14. Accessed May 9, 2025. https://www.pexels.com/photo/people-walking-near-buildings-1841568/.

Shabalala, Sizwe. 2019. "Photo of people walking near buildings." *Pexels*. April 16. Accessed May 9, 2025. https://www.pexels.com/photo/photo-of-people-walking-near-buildings-2464890/.

Singh, Yogendra. 2020. "People walking on street." *Pexels*. January 14. Accessed May 9, 2025. https://www.pexels.com/photo/people-walking-on-street-3444542/.

Trylski, Daniel. 2022. "Time Lapse Photography of Busy People Walking in the Entrance of an Arch." *Pexels*. October 13. Accessed May 9, 2025. https://www.pexels.com/photo/time-lapse-photography-of-busy-people-walking-in-the-entrance-of-an-arch-14018884/.