Project Deliverable D: Conceptual Design

Jack Croswell:	300356461
Fortune Effiong:	300356609
Louis Prince:	300383377
Zaid Lateef:	300424907
Theodore James Botman:	

October 12, 2024

Abstract:

Autonomous robots are the new frontier providing useful solutions to problems in a variety of sectors. However, despite autonomous robots' potential advantages, they pose an equally notable disadvantage. These disadvantages include digital dehumanisation, loss of meaningful human control, and lack of human judgement and understanding, and they majorly impact our relationship with technology. These notable downsides as well as guidance from the client provide the basis of the given design project.

Table of Contents

Introduction:

Project deliverable D's main purpose is to clearly define/refine ideas and conceptual aspects and organize them into given categories. Furthermore, it is the goal of this deliverable to develop a set of conceptual designs for our problem statement. As previously defined in past deliverables the problem statement is as follows: In an increasingly more technology-driven society, autonomous robots pose problems including; digital dehumanization, loss of meaningful human control, lack of human judgement and understanding, and impact on our relationship with technology. Furthermore, the basis of this deliverable is to use previous user benchmarking and technical benchmarking and the list of prioritized design criteria we have developed. By the end of this deliverable, the group should have analyzed and evaluated the concepts and choose the concepts or combination of concepts that they will continue to develop.

3 Defined Project Concepts:

3.1 CONCEPT 1: Dynamic Threat Detection System (Control)

3.1.1 Subsystem 1: Facial Recognition and Threat Assessment

The Robomaster's camera detects players and assesses their threat level based on movement and other types of behaviour. The more aggressive the movement, the more likely the system would target them. The subsystem leverages the Robomaster's camera to detect and identify players in the environment, using facial recognition combined with behavioural analysis to assess their threat level. The system evaluates movements and behavioural patterns, classifying players based on intensity and perceived aggression. The more aggressive a player is or the more sudden movements they possess, the higher their threat level is perceived by the robot, prompting it to target them more accurately.

Benefit: This illustrates how AI-driven or autonomous systems employ biased detection algorithms to select their targets. This opens a discussion about how biased data may lead to unfair targeting, especially in high-stakes environments where split-second decisions are more common.
Drawback: A concern is the misclassification of non-threatening individuals as high-risk based on the behaviour of the players. This would eventually result in false positives, where innocent players are targeted or treated as a big threat. This would further the idea of relying on AI in threat assessment and decision-making processes.
Metric:
Positive Facial Recognition: % of correctly identified players based on facial recognition.
False Positives: the rate at which the system mistakenly identifies non-threatening players as threats.
Monitor the incorrect threat assessments (eg. players walking slowly or standing still are being targeted)

Implementation into DJI:

Use the "when robot detects object" block to activate the facial recognition feature. You can set conditions to change the robot's behaviour based on what it detects.

For assessing threats, use the "if-then" block combined with "object detected" conditions (like aggressive movements). The robot's LED could turn red if a threat is detected, for example.

When an object is detected (face)

If movement is aggressive → Set LED to red / robot targets player

Else → Set LED to green

3.1.2 Subsystem 2: Self-Learning Behaviour

This subsystem forces the Robomaster to learn and evolve based on the player's interactions and responses. The robot master would essentially adapt its strategies over time, further modifying its behaviour to improve its accuracy when targeting. As the system collects more data, it becomes increasingly autonomous in its decision-making process. Each missed shot or attack would result in the data recalibrating and increasing its likeness in the next attacks. It learns from past mistakes and doesn't let it occur twice.

Benefit: Discuss how the Robomaster is unpredictable and adaptable. This would raise concerns about the robot's ability to operate outside of human control. Since it can very easily evolve, this reflects ethical issues surrounding autonomous systems' ability to learn and make decisions independently without further programming.
Drawback: As it continuously modifies behaviour, it may deviate from its original objectives. This could potentially lead to unintended outcomes and actions. This would explore the idea of accountability and ethical implications of machines that can evolve beyond their intended scope.
Metric:
Adaptation speed: the time it takes the robot to adjust behaviour. Track how many iterations it takes for the RoboMaster to modify its response when a player uses the same tactic (eg. hiding behind a shield, using the same decoy, etc.)
Learning Retention: The robot's ability to retain learned repeated behaviours over multiple sessions (or rounds of play). Track if the robot uses previously learned tactics when caught in similar scenarios.

Implementation into DJI:

Detect player actions:
→ If the behaviour is repeated multiple times (detected >2 times), update the robot's response
New round:
→ Use stored knowledge from previous rounds to adjust behaviour

3.1.3 Subsystem 3: Environmental Interaction

The players interact with the environment to either distract, hide, or manipulate the robot master's behaviour. They would need to use the objects around them as decoys and shields to protect themselves from the robot's unpredictable behaviour. These objects can be viewed as visual or behavioural cues made to mislead the RoboMaster, tricking the robot into misclassifying the situation. The players can throw or move objects to confuse the RoboMaster's tracking system. They can also use reflective surfaces to interfere with their sensors. They can also strategize to hide behind barriers that obscure their movements. This could effectively hide from the robot's line of sight and detection algorithms. This concept explores the robot's perception of environmental stimuli.

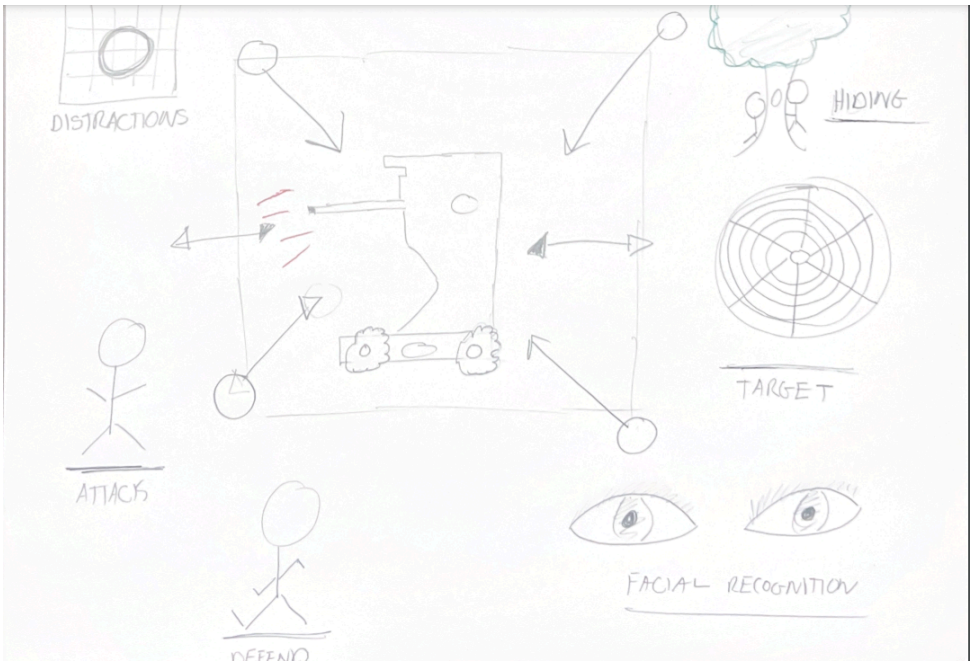
Benefit: Engages players to think about disrupting AI systems through environmental manipulation. It also forces players to think critically about how they can outsmart the robot. By forcing the players to manipulate the robot's actions, it discusses the importance of human ingenuity in overcoming autonomous systems.
Drawbacks: This would significantly limit interactivity provided by the environment. If the actions the players choose to take are restricted or too predictable, it may reduce the sense of realism within the simulation. This could decrease the feeling of a challenge and lead to repetitive gameplay.

Metric:
Player success rate: % of players that can effectively disrupt the RoboMaster's targeting based on environmental use.
AI Response time: how quickly and accurately the robot adapts to cues and player-created distractions
Variety of Strategies Used: the diversity of tactics the players choose to manipulate the robots targeting.

Implementation into DJI:

When the infrared sensor detects an object (decoy or player)
If object type = "decoy"
→ RoboMaster decreases threat priority
Else if object = "shield"
→ RoboMaster pauses targeting or repositions itself
Else if object type = "player"
→ RoboMaster resumes or increases threat assessment based on perceived movement (aggressive, defensive)

Sketch for Concept 1:



3.2 CONCEPT 2 - Evolving Attack and Defense (Ethics)

3.2.1 Subsystem 1: Adaptive Target Prioritization

The robot adjusts its attack priority based on the player's actions. Players performing defensive actions (eg. hiding, fleeing, defending) are deprioritised. The ones that are taking cooperative actions towards stopping the robot (eg. coordinating with others) are targeted first. The robot would essentially recognise when the players take or are planning to take action toward it.
The robot dynamically prioritises players who attempt to coordinate against it. The system focuses on players who are perceived as a greater threat, based on their actions. These actions include any acts of aggression towards the robot, wether it is the act of throwing something at it or moving towards it.

Once immediate threats are neutralised, the robot targets the players performing defensive action. The ones running away or hiding will be found and targeted.

Benefit: Explores how AI can adjust to human strategies and behaviours. This activity can show how AI can target potential offenders if they pose a threat. Conversely, this activity can insight some frustration in the players who practise defensive or evasive manoeuvres thinking they won't be targeted. This contributes to one of the main goals that the players feel like they lost regardless of their actions.
Drawback: might be difficult for the robot to accurately detect and identify the actions of the players, leading to confusion.
Metric: Frequency of target switches based on player actions. If a player runs towards the robot or holds up a sign signifying their intention to fight back, the robot will let out a sound effect. If a player runs away or hides, the robot will engage them after having targeted the formerly mentioned group of players.

Implementation in DJI:

Use the "when the infrared sensor detects object" block to identify players or objects in the robot's vicinity.

Set conditions using "if-then" blocks to adjust the robot's actions based on player behaviour. For example, if the player attacks, the robot can prioritize targeting them.

Example:

When the infrared sensor detects the player

If player action = attack → Robot targets the player first.

Else if player action = defend → Robot de-prioritizes the player.

3.2.2 Subsystem 2: Environmental Degradation

As the game progresses, the environment begins to degrade, simulating the collateral damage autonomous systems might cause in real-world scenarios. This can occur at scripted moments where the robot can knock down a stack of cups, representing a building, by running it over. Thus demonstrating the collateral damage.

Benefit: Highlights the unintended consequences of autonomous systems. It shows how the potential inaccuracy of automated killer robots can destroy innocent people.
Drawback: Excessive degradation may reduce strategy and lead to chaos. An increase in requirements for materials might effect the mobility of the game.
Metric: Amount of environmental damage over time. We can compare the targets that were meant to be destroyed to the peripheral damage. This ratio can be used to show the destructiveness of the robot.

Implementation in DJI:

Use a variable to represent the degradation level of the environment (e.g., "degradation_level").

Use "change [variable] by [number]" blocks to increase degradation after each round.

Example:

Every 30 seconds, change degradation_level by 1

If degradation_level > 5 → Robot moves slower or encounters obstacles in the environment.

3.2.3 Subsystem 3: Signal Disruption and Countermeasures
Players can attempt to disrupt the robot's signal, simulating real-world electronic warfare tactics.

We can simulated an EMP with a character or image shown to the robot, disabling it for a specified amount of time. This will give the player an opportunity to evade or escape temporarily.

Benefit: Introduces a layer of strategy by allowing players to interfere with the robot's operations.
Drawback: Signal disruption may become too dominant a strategy, reducing focus on ethical dilemmas.
Metric: The success rate of signal disruptions. The amount of successful evasions thanks to the device.

Implementation in DJI:

Use the "broadcast" block to simulate a signal disruption.

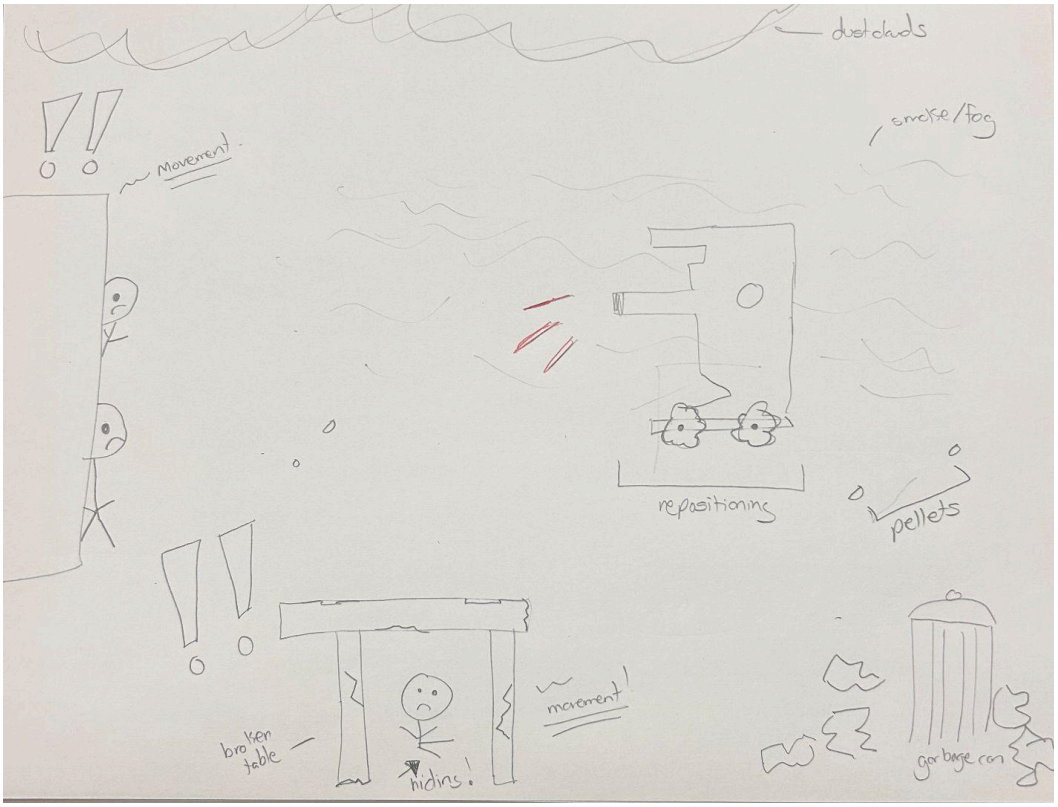
For example: when a signal disruption is broadcasted, the robot can stop its actions for a certain period.

Broadcast signal disruption

Wait for 5 seconds

Continue robot actions

Sketch for Concept 2:



3.3 CONCEPT 3: Human Control/Influence on Autonomous Robots (Hybrid)

3.3.1 Subsystem 1: Loss of Meaningful Human Control

With the use of autonomous robots in a variety of sectors not specifically in the defence sector, there is always the aspect of the loss of human control over robots. As robots get increasingly more and more advanced, there is a risk that those who programmed and developed said robots will have less and less control over what they have created. In general, the concern greatly affects the public as even with less advanced systems these autonomous robots may be able to be controlled by developers but not by the general public, therefore posing a major concern depending on the developer's intentions for their robot. As robots become more advanced, there is an increasing risk of losing meaningful human control. This subsystem demonstrates how, in the game, players gradually lose control over the robot as it becomes more autonomous.

Benefit: Illustrates the societal fear of losing control over advanced technologies.
Drawback: May be difficult to convey the robot's power with the limited capabilities of the robot.
Metric: Rate of control reduction over time.

Implementation into DJI:

Create a "variable" called "control_level" and decrease it over time using the "change [variable] by [number]" block.

Use "if-then" conditions to reduce player control as the game progresses. For example, if the control level drops below a certain point, the robot starts making decisions on its own.

Example:

Every 10 seconds, change control_level by -10

If control_level < 50 → Robot begins taking over control.

3.3.2 Subsystem 2: Programmer Power and Influence

If autonomous robots are found in the wrong hands they can be potentially controlled to do a wide variety of very bad things. This exemplifies the general concept that the general population would lack control over robots however a programmer does. Even though the nature of this idea that a programmer would create something for pure destruction seems dystopian it is not that far off considering the history of the wrong people coming into power. Tesla just released a human-like robot, what happens if this technology gets into the wrong hands? This subsystem explores how autonomous robots, if misused by their programmers, can be weaponized or controlled for harmful purposes.

Benefit: Demonstrates the potential dangers of putting too much power in the hands of programmers.
Drawback: Risk of players focusing on the narrative of a single company or person rather than the broader ethical issue.
Metric: Player engagement with ethical decision points.

Implementation into DJI:

Use the "broadcast" block to simulate when the programmer (AI) takes control. The broadcast event would trigger specific actions overriding the player's input.

Example:

Broadcast AI override

When AI override is received → Robot executes pre-programmed actions (e.g., "move forward," "fire blaster")

3.3.3 Subsystem 3: Government Regulation

An increase in autonomous robots could cause the technology to get into the wrong hands, furthermore, militaries may be able to implement the technology to create mass destruction. Therefore there is a need to control these robots through government regulation. However, the main concern is how governments could regulate these technologies. In the programming industry, there is a wide variety of methods for finding a singular solution to a problem. Therefore, how does one regulate something like programming and creating autonomous robots without there being loopholes? Thus there becomes the issue of lack of control over autonomous weapons robots in the government sector. This subsystem explores the difficulties of regulating autonomous technologies. Governments may struggle to keep up with the pace of innovation, creating loopholes that allow misuse.

Benefit: Highlights the regulatory challenges associated with autonomous weapons.
Drawback: Difficult to simulate government regulation in a gameplay scenario.
Metric: Number of loopholes or failures in regulation simulated.

Implementation into DJI:

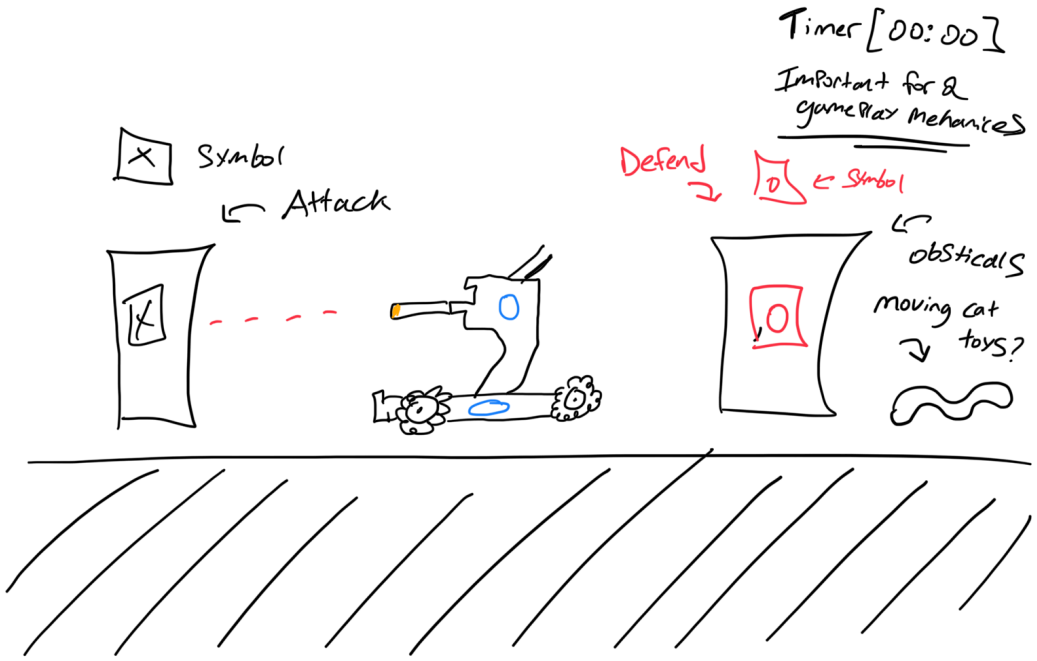
Use the "if-then" block to simulate when government regulations disable certain features of the robot (e.g., disabling the blaster after a certain number of rounds).

Example:

If round_number > 5 → Disable blaster.

If round_number > 5 → Robot no longer fires.

Sketch for Concept 3:



4. Global Concepts

4.1 Global Concept 1: Dynamic Control Focus

This idea focuses on how players interact with environmental elements and controls, with gameplay being primarily driven by the robot's threat detection and unpredictable behaviour. The main emphasis is on how players lose control as the robot's behaviour develops.

Advantages: Player interaction and control dynamics are highly emphasized.
Drawbacks: Might not effectively tackle the moral challenges inherent to the project.

4.2 Global Concept 2: Ethical Decision Emphasis

This idea emphasises providing ethical dilemmas and moral choices to players, with the robot serving as a neutral and developing system. The main objective is for players to experience the ethical impact of their decisions.

Advantages: Emphasises strong ethics and closely matches project objectives.
Drawbacks: Gameplay mechanics and control aspects receive reduced emphasis. It hard to measure the ethical impact, but the emotions of the user would be a telling indicator.

4.3 Global Concept 3: Balanced Hybrid Design

This design combines both ethical dilemmas and player control dynamics, blending the best of the previous concepts. Players face moral decisions while gradually losing control of the robot, resulting in a complex and immersive experience.

Advantages: Well-balanced, engaging gameplay with meaningful moral decisions. Will highlight the powerlessness felt by the players as the robot learns and defeats them.
Drawbacks: May require careful tuning to prevent overwhelming players.

5. Evaluation and Selection Matrix (Based on Deliverable C)

Criteria	Concept 1 (Control)	Concept 2 (Ethics)	Concept 3 (Hybrid)
Interactive Gameplay	High	Medium	High
Progressive Loss of Control	Medium	Low	High
Moral and Ethical Dilemmas	Low	High	Medium
Post-Game Reflection	Low	High	High
Accessibility	High	Medium	High

Interactive Gameplay:

Concepts 1 (Control) and 3 (Hybrid) received high scores because they emphasize engaging the player through dynamic interactions with the robot. Concept 2 (Ethics) focused more on ethical decision-making compared to Concept 1, resulting in a moderate score.

Progressive Loss of Control:

Concept 3 received a high score due to its gradual decrease in player control as the robot assumes more responsibility, a crucial element of the game's design. Concept 1 displayed some loss of control, though not in a continuous manner, resulting in a moderate rating. Concept 2 had a decreased emphasis on this aspect, leading to a lower rating.

Moral and Ethical Dilemmas:

Concepts 2 and 3, Ethics and Hybrid, received high scores by centring on providing players with challenging moral decisions that mirror ethical dilemmas in the real world. Concept 1, on the other hand, focused more on control mechanisms, achieving a lower score in this aspect.

Post-Game Reflection:

Both Concept 2 and Concept 3 were highly rated for encouraging players to contemplate the ethical consequences of their actions post-game. Concept 1 received a lower score because it did not prioritize post-game reflection as much.

Accessibility:

Concepts 1 and 3 received high scores because of their easy-to-understand gameplay mechanics, which appeal to a wide variety of players. The increased ethical challenges and decision-making aspects of Concept 2 made it slightly less accessible, resulting in a moderate rating.

Based on this matrix, **Concept 3** was selected for further development. It offers a compelling combination of user interaction and ethical challenges while maintaining a focus on gameplay mechanics.

6. Conclusion and Future Work

The hybrid design chosen offers a well-rounded and engaging experience by blending player control mechanics with impactful ethical choices. These choices will be used to show that AI killer bots can be inaccurate, destructive and biassed. Proceeding, the upcoming stage will concentrate on improving the subsystems, specifically the decision-making logic and control mechanisms, to guarantee their smooth integration. Client feedback will be included in future development to ensure that the design continues to meet project objectives. This feedback will help us fine-tune our systems relating to specific capabilities of the robot. Moreover, the feedback will help us narrow our scope on the types of ethical decisions to be made by the players in order to properly highlight the dangers of killer robots.

7. References

DJI. "RoboMaster S1 Specifications." DJI, www.dji.com/robomaster-s1/specs.

Robin the Robot. “The Progress Network,” May 9, 2022.

Terminator Ethics: Should We Ban Killer Robots? Ethics & International Affairs, 2015.