

# Playtesting (and AI Approaches)

---

# Playtesting (and AI Approaches)

---

By

**Chrysa Bika, Daniel Dyrda,  
Martin Schacherbauer,  
Johanna Pirker**

TUM, Munich, Germany

Course

**AI in Games**

TUM

Version

**1.0**

# Content

---

1. Introduction to Playtesting (and User Studies)
2. Measuring Player Experience
3. AI for Playtesting (Case Studies by Daniel Dyrda)

# Player Experience

---

# Player Experience

---

- The Player experience arises from interaction with the game—by **playing the game**.
- A game is not really about its systems or its assets. A **game is about the experience it makes possible through play**.
- The indirect nature of PX is identified as **a second-order design problem** [257] “... Meaning that designers are communicating with players indirectly through their games” [289,p. 165].
  - The designer controls only the first-order elements—the mechanics, rules, and systems—while the player experience emerges indirectly as a second-order effect.

→ **With playtesting, we want to test hypotheses about the player experience.**

# Playtesting

---

# Playtesting

---

Playtesting is the process of **observing other players while playing** your game. During this process, you should **not interrupt or give tips**. Instead, sit in silence and try to not disturb the player.

The **goal** of playtesting is to **evaluate whether the assumptions** you made about your game, or a specific level, are true or not. The goals need to be **clearly defined**, so you can determine whether your current state of game accomplishes the goals.

**Playtesters** should represent your game's **target audience**, but including other groups of people that are not in your target audience can benefit your game's quality in the end. Valve also uses the different view perspectives of different groups of players to test specific parts of games – such as FPS-players for a boss fight to evaluate its intensity.

# How Valve Playtests

---

The developers at Valve always set **goals** for a level or game. Then they **design** with that goal, followed by **playtesting** whether the goals are met. Then, they **iterate** and do the whole process over again with the learnings from the playtest.

- They **repeat this process** until it is not “excruciatingly painful to watch the playtests”.
- They see their **design as a hypotheses** and use playtests as an experiment to **validate this hypotheses**.



Valve's “Secret Weapon” – Game Maker’s Toolkit (YouTube)

# How Valve Playtests

---

"Playtesting is simply **watching people play** through a chunk of the game - sometimes with a questionnaire or interview at the end - and then **using what you see and hear to drive changes** in the game's design." ~1:40

Playtest feedback is just data → Game designers need to know how to interpret, filter, and apply that data to their own game.



Valve's "Secret Weapon" – Game Maker's Toolkit (YouTube)

# Six Key Takeaways from Valve

---

- **Playtest early** – Art, sound, etc. does not matter. What you want to test is your game mechanic!
- **Test often** – By testing more frequently, you can gather huge amounts of data to draw your conclusions from
- **Shut up and watch** – The “obvious” answer you designed might not be so obvious after all...
- **Designers should do playtests** – The designers know their vision best and what conclusions need to be drawn from player’s feedback. Also, it motivates to see problems and that you can fix them!
- **Get the right people** – The more variance in your playtesters, the more varied your feedback will be. Also, by having more people giving feedback, outliers can be determined more easily.
- **Challenge your assumptions** – If 99/100 people do not agree with your choice, there is a high chance of them being correct and your assumptions for making that choice being biased, flawed, or false.

## More Perspectives

---

**Totten:** "Playtesting - evaluating a level by having people play it many times to test for experiential and technical functionality - is the standard way of ensuring that levels are good." (Totten, 2019, p.xxvi) "As playtesters try your levels throughout your process, document what went well with each prototype and what could be changed or improved. Change your design and have someone playtest your next prototype" (Totten, 2019, p.86)

**TychoBold:** "Playtest early and often. It's easy for you to become blind to the problems in your level. Playtests should drive the iteration, not the other way around. It's easy to get into the mindset of "Just wanting to polish it a little bit more before a playtest". Playtesting is to learn what you need to change. The later and less frequently you playtest, the more expensive the needed changes could become." (@TychoBold (K., Alex), 2020, 2020, p.15)

# **Doing It Right – (A Minimal Setup)**

---

## Doing It Right – (A Minimal Setup)

---

- Ethics
- Informed Consent
- Demographics
- Privacy and Data Protection
- Special Considerations for Sensitive Content

# Ethics

---

Always adhere to the ethical standards of Germany and the European Union! If you want to publish your findings in a scientific publication, you need the approval for your user study by an ethics board.

- Ethics Commission of TUM: <https://www.tum.de/en/about-tum/organization/ethics-committee>
- Ethics Commission Basics TUM: <https://www.ek-med-muenchen.de/index.php>
- Ethics Commission of faculty 16 of LMU: <https://www.ek.mathematik-informatik-statistik.lmu.de>

# Informed Consent

---

## Step 1: Inform the User

- **Explain the purpose:** Participants should understand what the playtest involves — what they'll do, how long it will take, and why data is being collected.
- **Voluntary participation:** No one should feel pressured to take part. Participants should be free to withdraw at any time without penalty.
- **Age-appropriate consent:** For minors, get parental or guardian consent, and ensure the content is suitable for their age group. (Working with minors requires the approval by an ethics board!)

# Informed Consent

---

## Step 2: Obtain Informed Consent

Obtain “informed consent” from the test person stating that...

- ... they have read the task description.
- ... they have been informed that they can quit.
- ... they agree to being recorded/tracked (if that’s the case.)
- ... that statements from interviews can be quoted.
- ... that test organizers are allowed to look at the data.

# Demographics

---

Collect anonymized demographics of the users.

- This is mandatory if you plan to use the data in scientific publications!

# Privacy and Data Protection

---

Anonymize all user-related data immediately.

- **Anonymize data:** Remove names and identifying details from feedback and recordings.
- **Secure storage:** Store gameplay recordings, surveys, or personal data only on trusted devices.
- **Transparency:** Clearly explain what data is being collected (e.g., playtime, behavior, reactions) and how it will be used.

## Special Considerations for Sensitive Content

---

If the game includes mature or potentially triggering material (e.g., violence, trauma, social issues):

- Provide **content warnings**.
- Allow **opt-out options**.
- Offer **debriefing sessions** to discuss reactions afterward.

# Ways to Measure Player Experience

---

# Self-Reported Data

---

Self-reports can take various forms, such as

- **interviews and questionnaires** conducted after play or
- **thinking-aloud** methods during play.

While allowing the estimation of PX, **self-reported data is limited** as it is **subjective, influenced by environmental factors and formalization**, and **lacks precise linkage** to specific play events when recorded retrospectively.

## Self-Reported Data

---

Standardized questionnaires such as the **Player Experience Inventory**, **PENS**, and **GEQ** assess specific PX dimensions, with additional tools reviewed by Wiemeyer et al.

Many questionnaires are not or only partially empirically validated, and they are constrained in their capacity to capture holistic experiences.

# Behavioral Data

---

PX can also be estimated using **more objective measurements** of what the player is doing. **Personal experience often manifests through observable behaviors** such as laughing, smiling, or frowning.

- Several studies have **proposed methods for real-time measurement** using behavioral data derived from the game and the physical behavior of players.
- This analysis can include any player-initiated behavior ranging from low-level data, such as button presses and input vectors, to in-game interaction data with regard to the game mechanics and the game state.

While these approaches yield valuable insights into player behavior and, indirectly, PX, they **often fall short of capturing emotional and affective states.**

# Psychophysiology

---

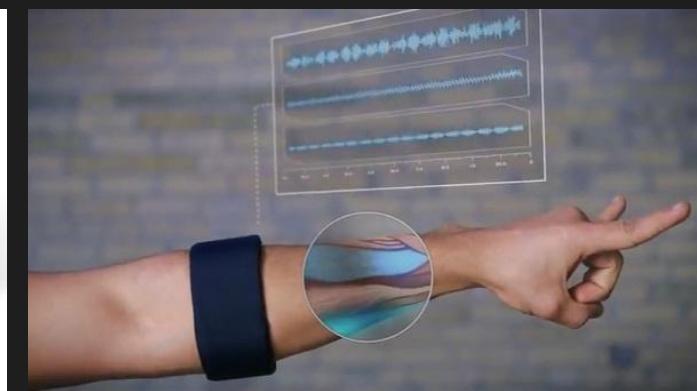
**Physiological measurements**, including skin conductance, heart rate, blood pressure, eye movement, and brain activity, **offer a means to uncover** hidden changes in a **player's psychological state**.

- **Psychophysiology** examines the **interplay between psychological processes and their physiological correlates**. Neuroimaging methods, such as electroencephalography (EEG) and functional magnetic resonance imaging (fMRI), enable **real-time estimation of PX** by detecting player state changes.
- These metrics have the **potential to reveal emotional responses** that might remain undetectable using traditional methods. **Observable PX qualities include challenge, stress, and mental functions**, including attention, memory, and decision-making.
- Application is **constrained by hardware limitations, cost, and the reliability of deriving PX quality** from physiological measurements.

# Bio-Feedback

---

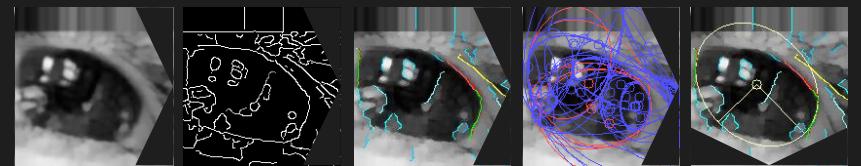
- Electroencephalography (EEG)
- Muscle measurements
- Electrocardiography (ECG)



# Eye-Tracking

---

- Detect the user's eye position
- Determine where user looks at while playing
  - Note: Huge difference between professional players and casual players



# Beyond Playtests: User Studies

---

## Beyond Playtests: User Studies

---

If you want to conduct **user studies for a scientific context**, i.e. publishing your study in a scientific publications, there's more!

# Evaluation Approaches – A Brief Overview

---

## Phases and Strategies

### Phase 1: before development

- Expert reviews
- User surveys

### Phase 2: during development

- Usability testing and laboratories
- User surveys

### Phase 3: after development

- Acceptance tests
- Evaluation during active use
- User surveys

## Methods

### Phase 1

- Cognitive walkthrough
- Heuristic evaluation

### Phase 2

- Formative evaluation
- Interviews and demos
- Questionnaires
- Summative evaluation

### Phase 3

- Questionnaires
- Interviews and demos

# Statistics

---

Statistical Consulting at TUM: <https://www.math.cit.tum.de/en/math/department/statistical-consulting/>  
with consulting service, research data management, statistical courses, literature and statistical software.

## More Information

---

In this course, we want to focus on classical Playtesting. For more information about user studies, we refer to the course **“3D User Interfaces”** and the book **“Research Methods in Human-Computer Interaction”**.

### RESEARCH METHODS IN HUMAN-COMPUTER INTERACTION

Second Edition



Jonathan Lazar | Jinjuan Heidi Feng | Harry Hochheiser

# AI for Playtesting

---

# Player Experience

---

- The Player experience arises from interaction with the game—by **playing the game**.
- A game is not really about its systems or its assets. A **game is about the experience it makes possible through play**.
- The indirect nature of PX is identified as **a second-order design problem** [257] “... Meaning that designers are communicating with players indirectly through their games” [289,p. 165].
  - The designer controls only the first-order elements—the mechanics, rules, and systems—while the player experience emerges indirectly as a second-order effect.

→ **With playtesting, we want to test hypotheses about the player experience.**

# AI for Playtesting

---

How can AI support us with the playtesting process?

# AI for Playtesting

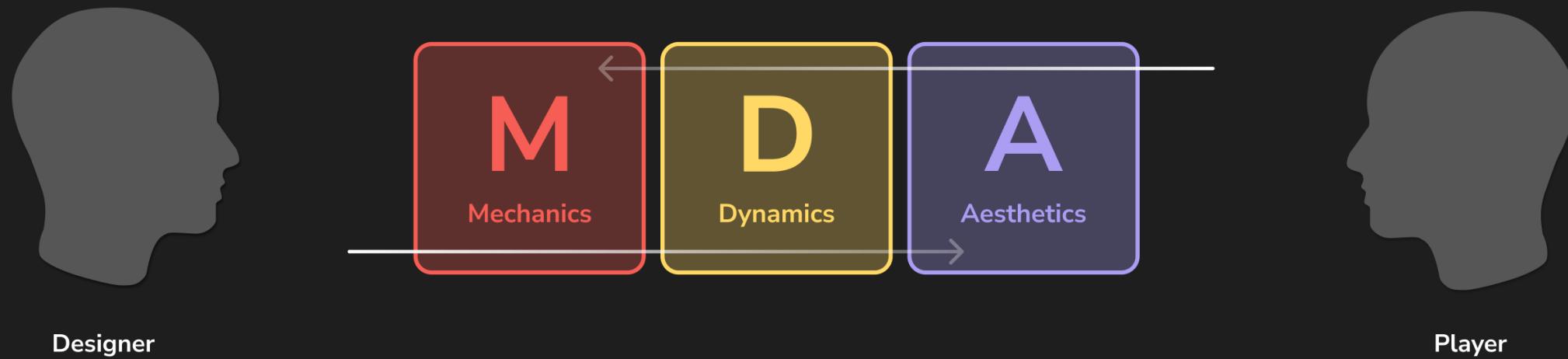
---

**How can AI support us with the playtesting process?**

Perspective: You do not need to be an AI researcher to apply AI methods in the playtesting/design evaluation phase.

# Playtesting

---



# The MDA Framework

---

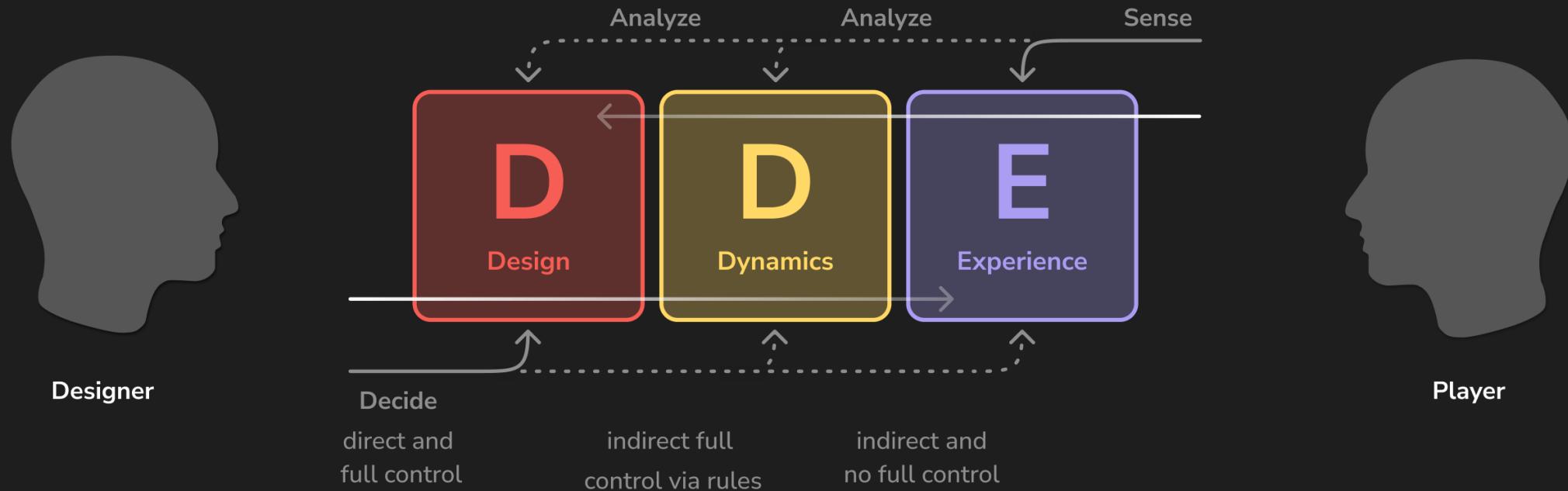
A perspective often applied to understand what happens at the game interface—the meeting point between the human and the game—is the Mechanics-Dynamics-Aesthetics (MDA) framework [161]. It conceptualizes games as a chain of relationships between Mechanics, Dynamics, and Aesthetics.

- **Mechanics** are the designed rules and components of the game artifact;
- **Dynamics** describe the system behavior that emerges when these mechanics interact with player input and each other over time; and
- **Aesthetics** refer to the desirable emotional responses and experiences that arise for the player when playing the game.

From the designer's perspective, the process flows from Mechanics to Dynamics to Aesthetics, while from the player's perspective, it is experienced in reverse—from Aesthetics to Dynamics to Mechanics—as players engage with and interpret the designed systems through play.

# Playtesting

---



# The DDE Framework

---

Building on the MDA framework, the Design-Dynamics-Experience (DDE) framework [306] extends and refines this relationship, addressing some of MDA's limitations while retaining its core structure.

Both frameworks share a key insight: player emotions and experiences are indirect effects—emergent outcomes of systems that designers construct but do not directly control.

- In DDE, **design** encompasses everything the designer can deliberately shape, extending beyond what MDA defined as mechanics.
- **Dynamics** remain the emerging system behavior when features interact with player input and each other over time, producing the evolving behavior of the game system.
- Finally, DDE replaces MDA's term aesthetics with **experience**, clarifying that the focus lies not only on emotional responses but on the full range of cognitive, affective, and embodied experiences that emerge through play.

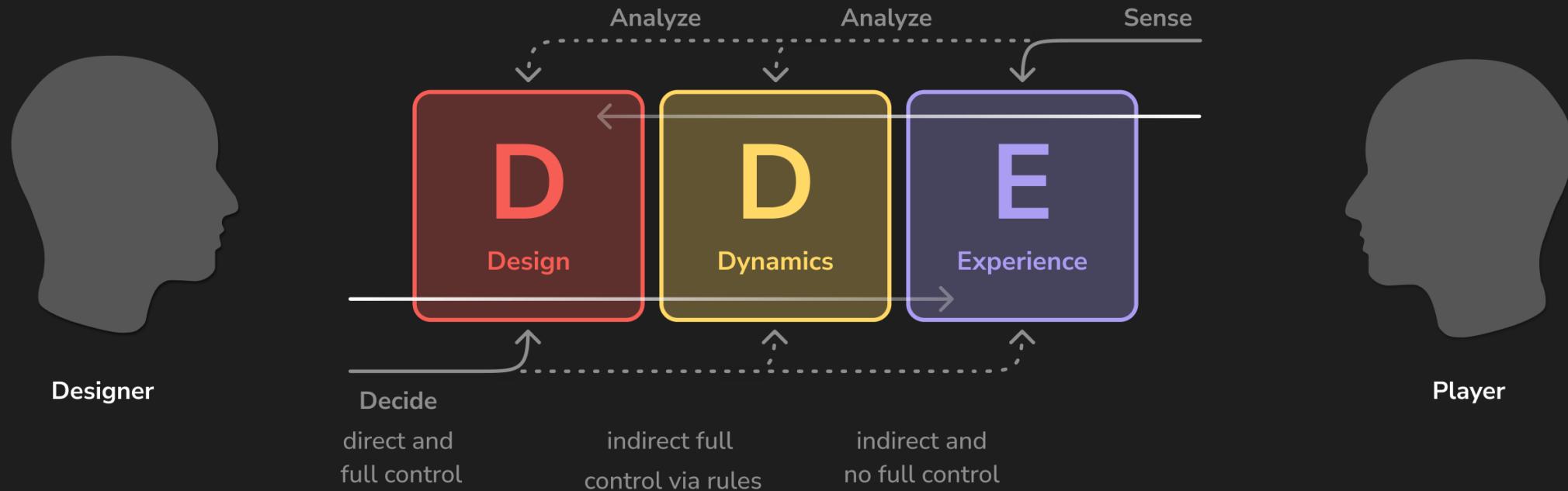
## The DDE Framework

---

While the framework also agrees with the sequencing of both perspectives' process flow, the framework adaptation makes clear that this sequence is not linear, but should be understood in a parallel fashion. "From the designer's perspective, this non-linear sequence is an iterative process, where each decision on every level can alter the whole game. From the player's perspective, we can further see this non-linearity play out in the experience of a game. The player senses the game world and analyzes the underlying mechanics at the same time, while confronting and experiencing every aspect of the design." [306, p. 12]

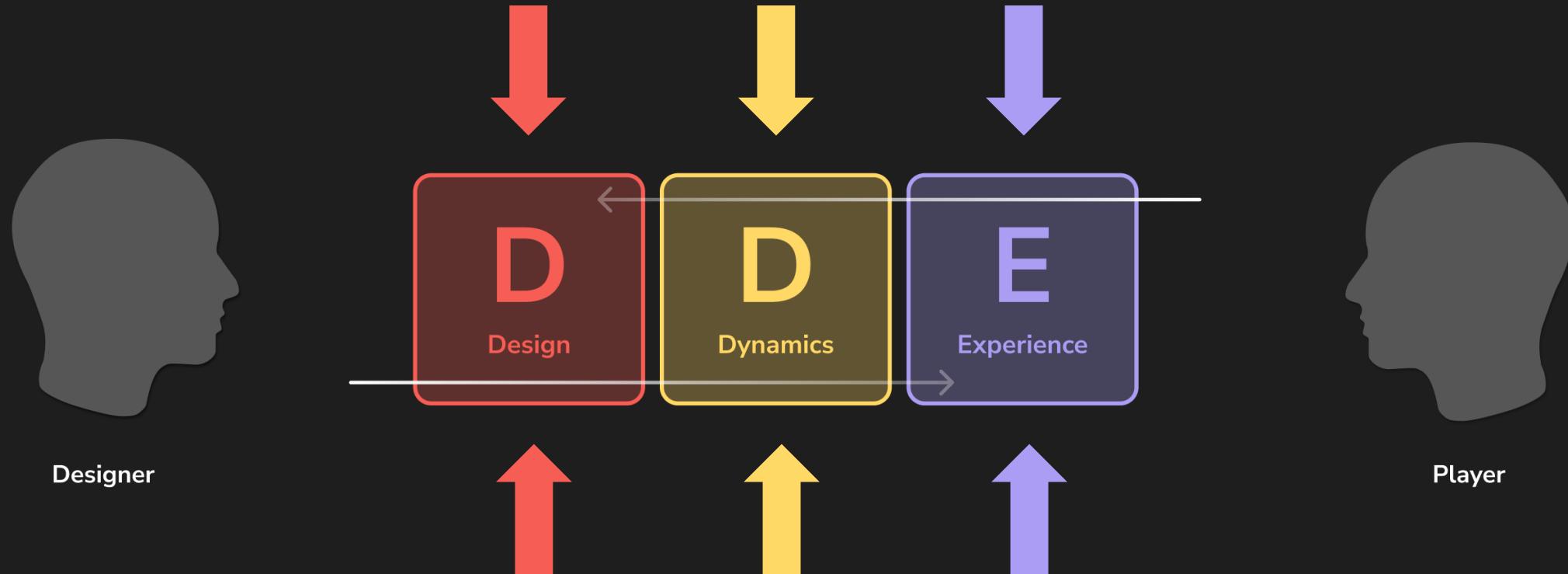
# Playtesting

---



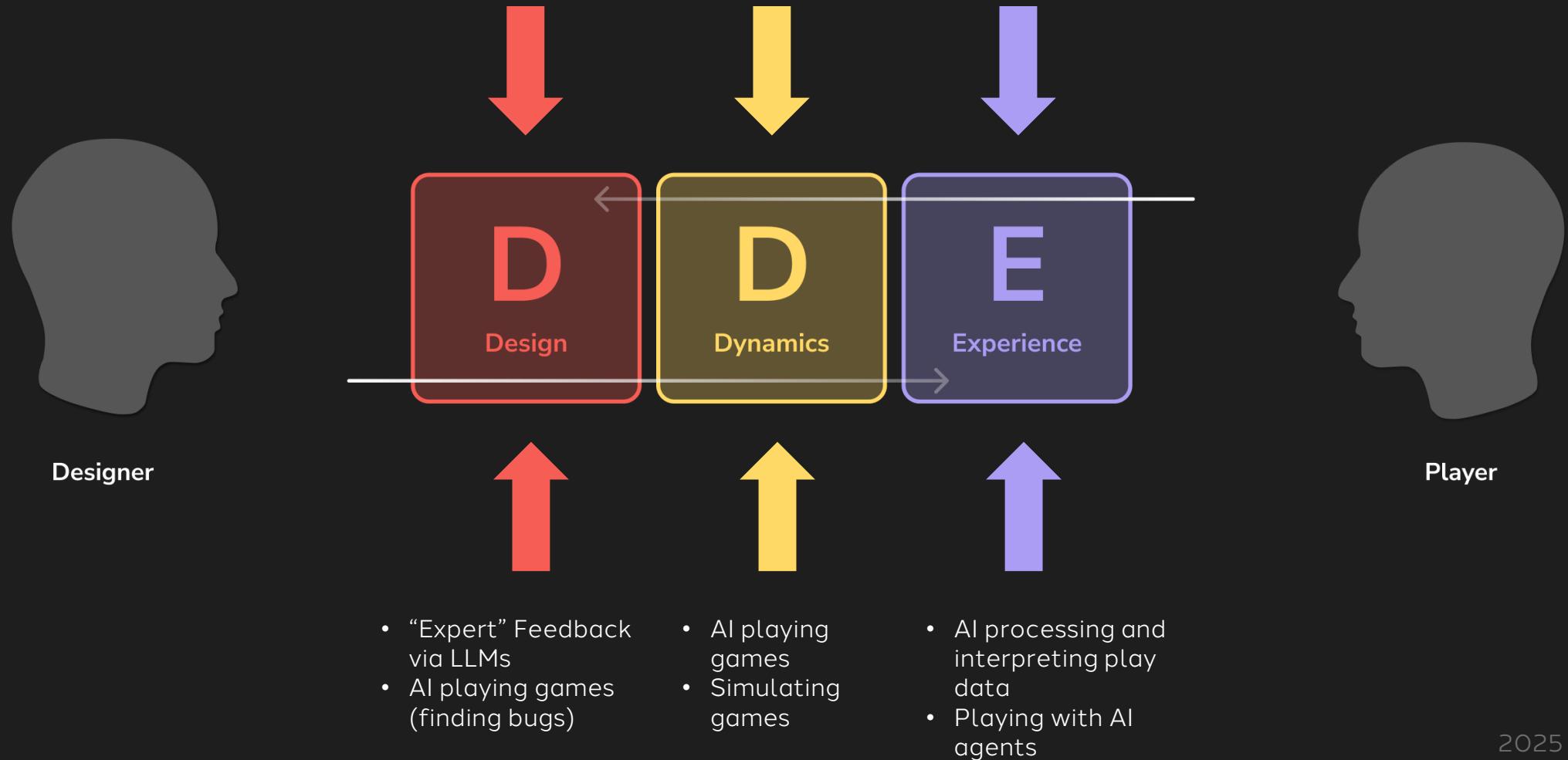
## Play-Testing – How can we support Play-Testing with AI?

---



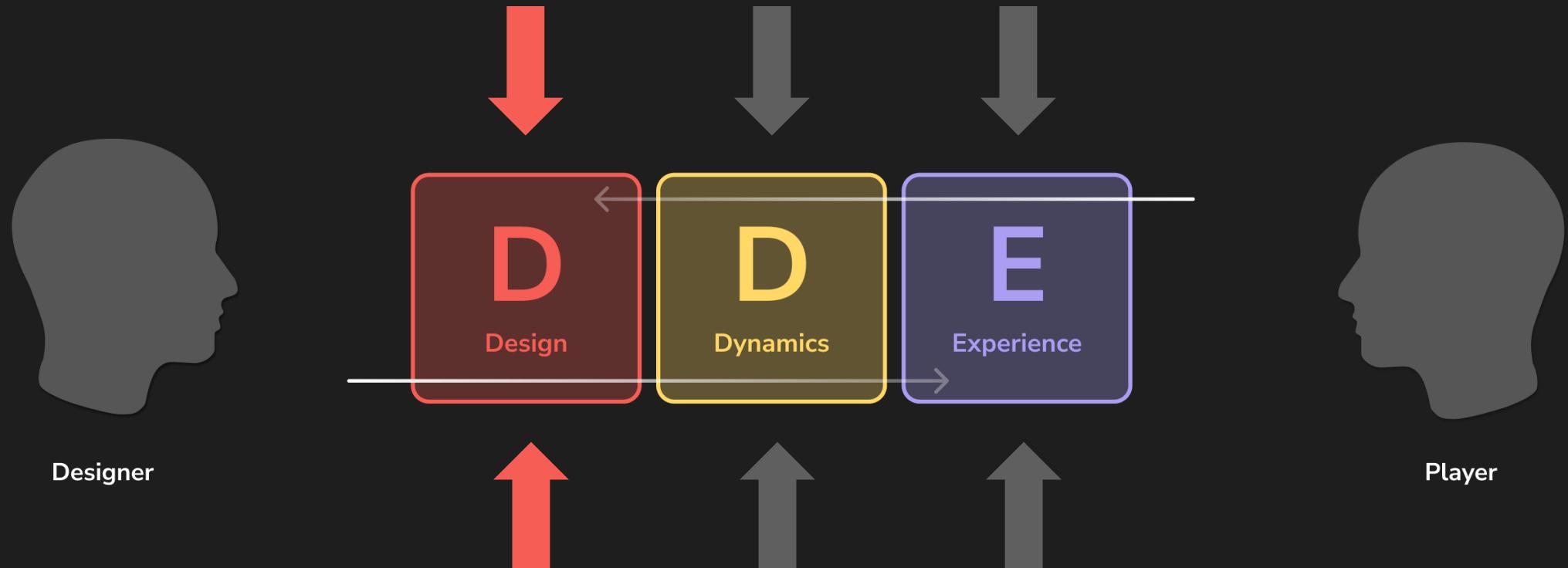
# Play-Testing – AI Approaches

---



# Play-Testing – AI Approaches

---



- “Expert” Feedback via LLMs
- AI playing games (finding bugs)

- AI playing games
- Simulating games

- AI processing and interpreting play data
- Playing with AI agents

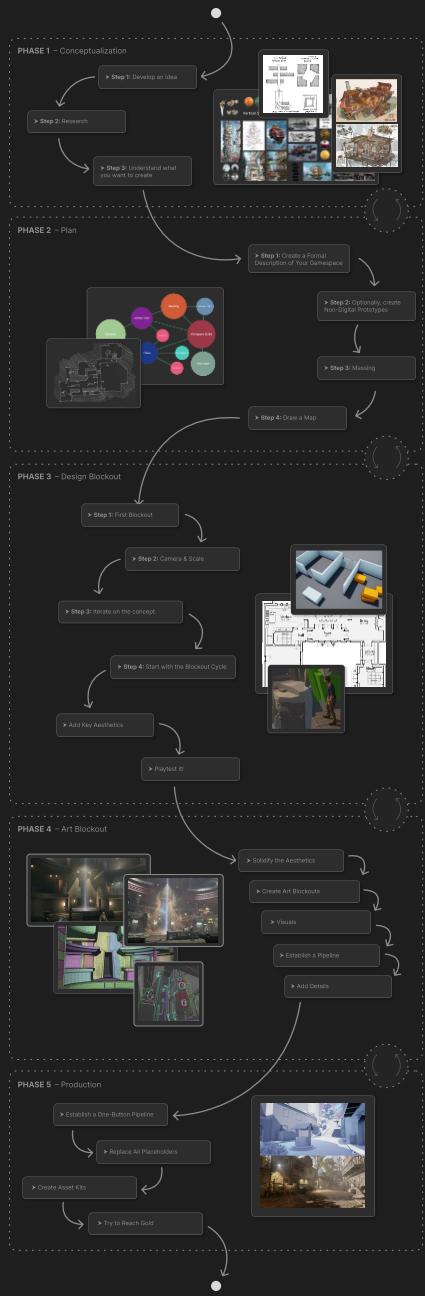
# **Case Study: Design Evaluation with LLMs**

---

# Using LLMs to Evaluate Game Ideas

---

Idea: Use an LLM to evaluate game concepts based on ten different aspects of a game concept (e.g., player experience, place, story and narrative)

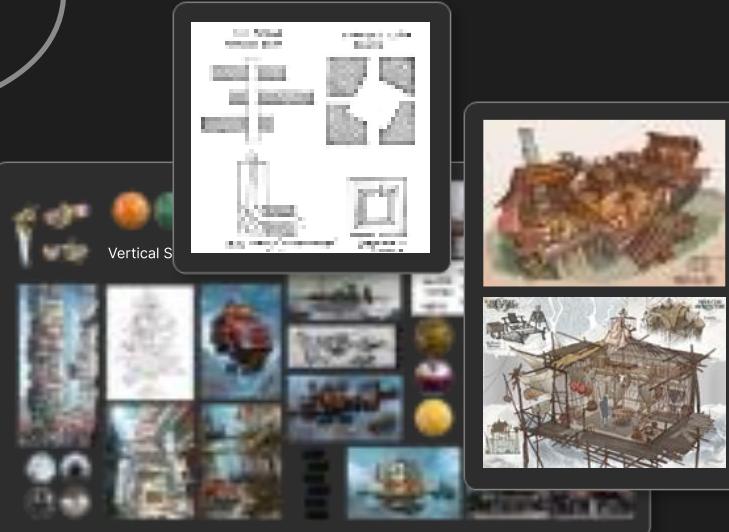


## PHASE 1 – Conceptualization

► Step 1: Develop an Idea

► Step 2: Research

► Step 3: Understand what you want to create



## PHASE 2 – Plan

► Step 1: Create a Formal Description of Your Gamespace

► Step 2: Optionally, create Non-Digital Prototypes

## ► Step 1: Develop an Idea

---

Before you can work on something, you need an idea. When you think about something that is experienced by a player when playing your game, we call this a **Player Experience Idea**. This idea can be related to an entire game, a level, or a simple encounter, such as a boss fight. When you have an idea for a player experience, you should be able to answer the following design questions.

Document your answers.<sup>[1]</sup>

◆ Player Experience >

◆ Theme >

◆ Gameplay >

◆ Place >

◆ Unique Features >

◆ Story & Narrative >

◆ Goals, Challenges & Rewards >

◆ Art Direction >

◆ Purpose >

◆ Opportunities & Risks >

## ❖ Gameplay ▾

Describe the core gameplay.

- Try to find 3-5 verbs that describe the gameplay experience.
- Describe what Core Mechanics are relevant for your idea.
- Describe what the player does by formulating a 30 Seconds of Gameplay.
- If you have special Level Core Mechanics make this very clear.

## ❖ Place >

## ❖ Unique Features >

## ❖ Story & Narrative ▾

Come up with a rough story. Think about how the player will experience this story: how does the game or level tell this story? Think about storytelling methods, such as environmental storytelling, gameplay, cutscenes, narrators, dialogues, ...

- What is the story of the environment or game?
  - Write a short description of the environment from the perspective of a character that lives here or someone who has been here already ([Galuzin, 2016](#))
- What happened here before the player arrived?
- What will the player experience here?
  - The arrival of the player.
  - How?
    - How did the player arrive in this location? Link documents of other locations here.
    - What were the events that brought the player here?
  - Why? The player goals

# Using LLMs to Evaluate Game Ideas

---

**First Phase** – Determine the optimal LLM model

- Generate game summary by LLM
- LLM gets information about aspects of game concepts
- Determine which AI model to use based on human evaluation

**Second Phase** – Use best-performing model from phase 1

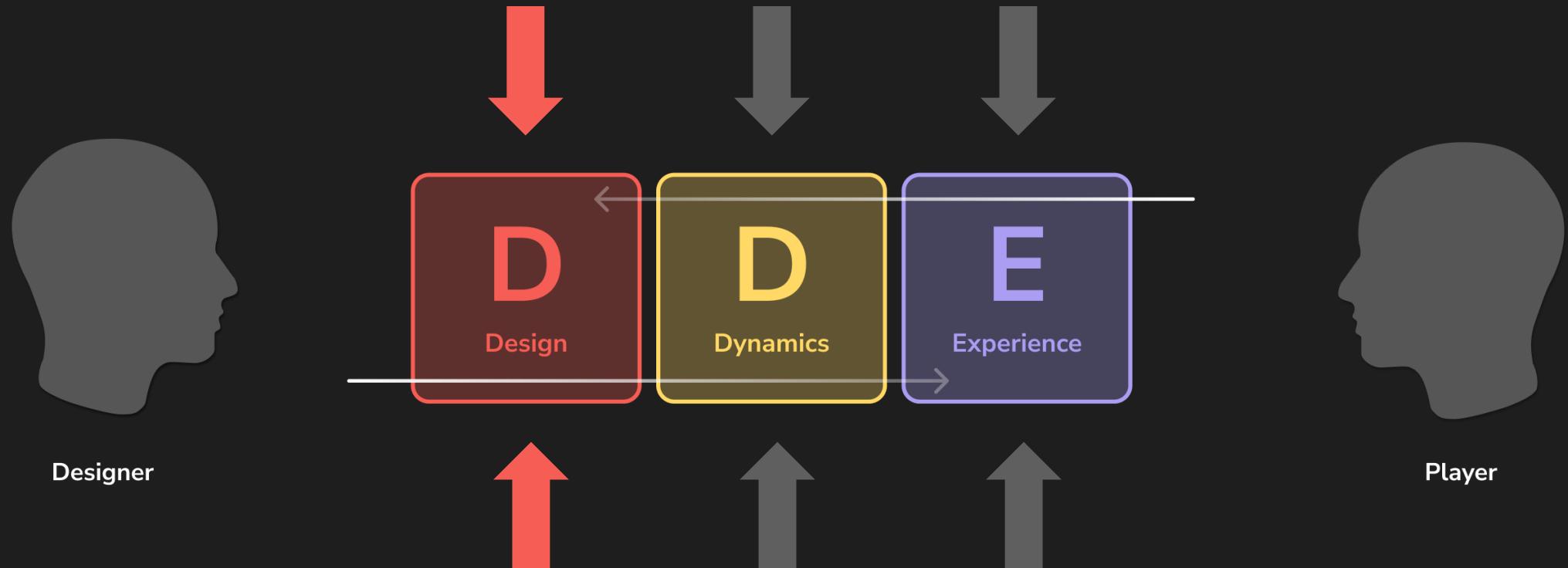
- User study
  - Participants submit a game concept
  - LLM evaluates the game concept based on the ten aspects
  - Participants receive evaluation and fill out questionnaire

## Takeaways

- Longer game concept → More general answer of LLM
- Prompt “engineering” main factor to improve reliability

# Play-Testing – AI Approaches

---



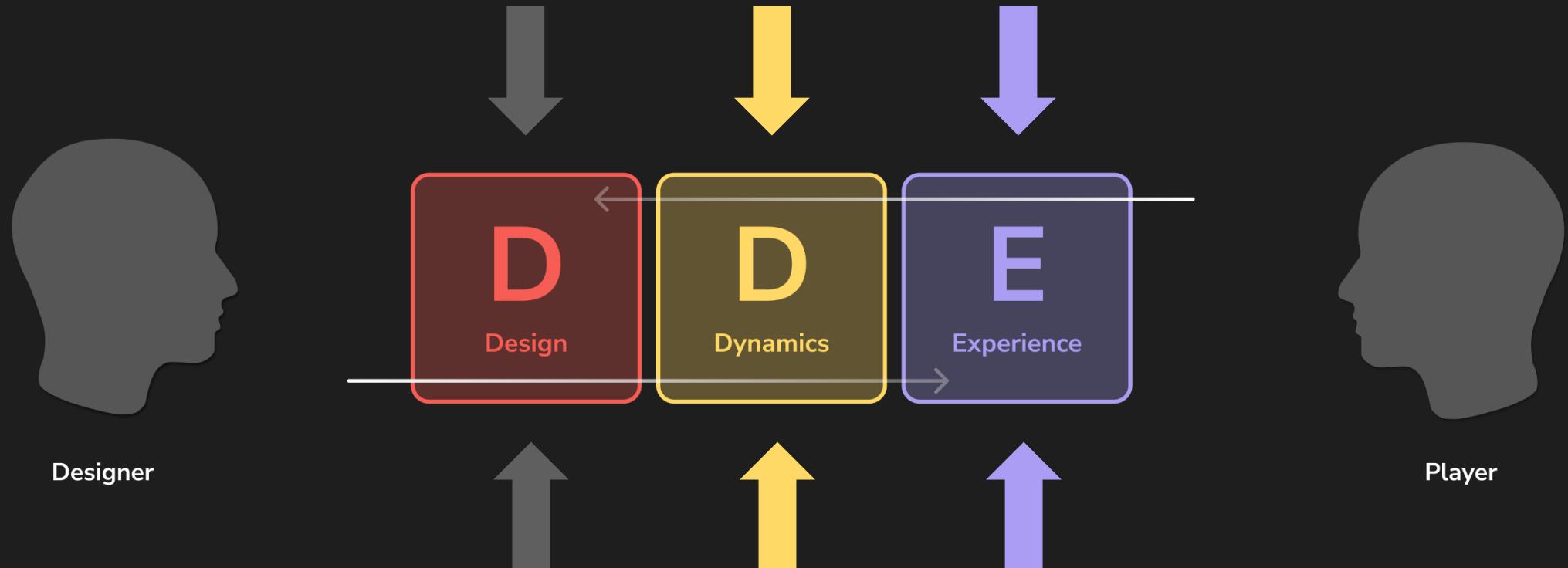
- “Expert” Feedback via LLMs
- AI playing games (finding bugs)

- AI playing games
- Simulating games

- AI processing and interpreting play data
- Playing with AI agents

# Play-Testing – AI Approaches

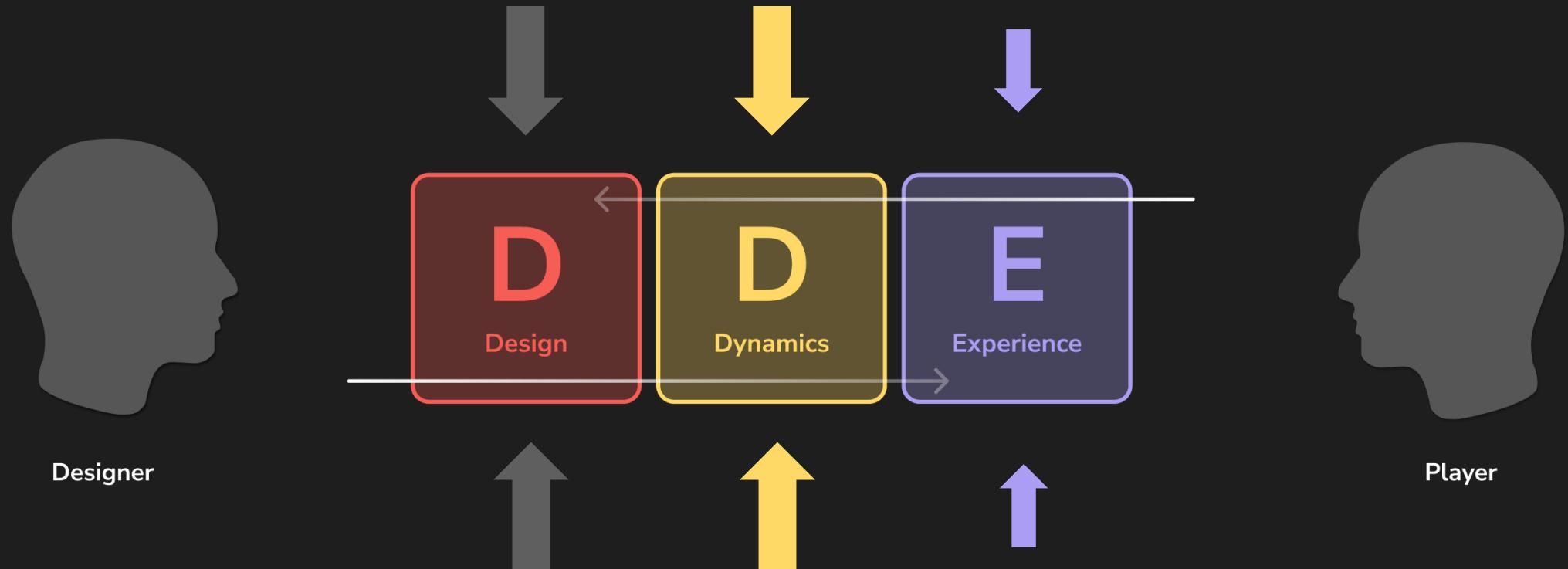
---



- “Expert” Feedback via LLMs
- AI playing games (finding bugs)
- AI playing games
- Simulating games
- AI processing and interpreting play data
- Playing with AI agents

# Play-Testing – AI Approaches

---



Designer

Player

- “Expert” Feedback via LLMs
- AI playing games (finding bugs)
- AI playing games
- Simulating games
- AI processing and interpreting play data
- Playing with AI agents

# **Case Study: Agent-Based Modelling**

---

# Agent-Based Modelling

---

Agent-Based Modelling (ABM) is a computational approach that simulates the actions and interactions of individual autonomous agents to explore how complex system-level patterns emerge from simple local rules.

- **Agents:** Autonomous units with distinct behaviors and goals.
- **Environment:** The space (physical or abstract) where agents interact.
- **Emergence:** System-wide patterns arise from lower-level interactions.
- **Adaptation:** Agents can learn, evolve, or respond to feedback over time.

# AI agents can “play” games

---

What do we mean by this?

- Simulation vs. actual “playing” (using mechanics)
- Agent for ML based approaches
- Are games always completely observable?

→ Case Study: Rule-Based Agent Simulation for World of Warcraft

→ Case Study: ML Agents for Testing Level Design

	Player	Non-Player
Win	<p><b>Use Cases</b></p> <p>Games as AI Testbeds AI that Challenges Players <i>RL Benchmarking</i> <i>General Game Playing</i></p> <p><b>Examples</b></p> <p>TD-Gammon, Chinook, Deep Blue, Watson, AlphaGo, AlphaStar</p>	<p><b>Use Cases</b></p> <p>Game Balancing Playing Roles that Humans Would not (Want to) Play</p> <p><b>Examples</b></p> <p>Rubber Banding in Racing Games</p>
Experience	<p><b>Use Cases</b></p> <p><i>Content Evaluation,</i> <i>Multiplayer Team Bots,</i> <i>Ghosts and Personas,</i> <i>Game Testing, Demo Mode</i></p> <p><b>Examples</b></p> <p>Game Turing Tests (2kBot Prize) Drivatar in Forza Series</p>	<p><b>Use Cases</b></p> <p>Believable &amp; Human-like Agents, <i>Adversarial NPCs,</i> <i>Cooperative NPCs</i></p> <p><b>Examples</b></p> <p>Ellie in <i>The Last of Us</i>, Alyx in <i>Half Life 2</i>, Elites in <i>Halo Series</i></p>

	Player	Non-Player
Win	<p><b>Use Cases</b></p> <p>Games as AI Testbeds AI that Challenges Players <i>RL Benchmarking</i> <i>General Game Playing</i></p> <p><b>Examples</b></p> <p>TD-Gammon, Chinook, Deep Blue, Watson, AlphaGo, AlphaStar</p>	<p><b>Use Cases</b></p> <p>Game Balancing Playing Roles that Humans Would not (Want to) Play</p> <p><b>Examples</b></p> <p>Rubber Banding in Racing Games</p>
Experience	<p><b>Use Cases</b></p> <p><i>Content Evaluation,</i> <i>Multiplayer Team Bots,</i> <i>Ghosts and Personas,</i> <i>Game Testing, Demo Mode</i></p> <p><b>Examples</b></p> <p>Game Turing Tests (2kBot Prize) Drivatar in Forza Series</p>	<p><b>Use Cases</b></p> <p>Believable &amp; Human-like Agents, <i>Adversarial NPCs,</i> <i>Cooperative NPCs</i></p> <p><b>Examples</b></p> <p>Ellie in <i>The Last of Us</i>, Alyx in <i>Half Life 2</i>, Elites in <i>Halo Series</i></p>



# **Case Study: Rule-Based Agent Simulation**

---



# Rule-Based Agent Simulation

---

- **Goal:** Compare different classes in regards to certain values (such as DPS)
  - We can use simple rule-based agents to simulate DPS
    - Require rules for each class OR rules that apply for multiple classes
  - Simulated DPS value → Can be used for balancing
- **Problem:** Data gathering can be difficult (similar skill level required, many players)
- **Examples:** Raidbots, SimulationCraft

## Buff Uptime



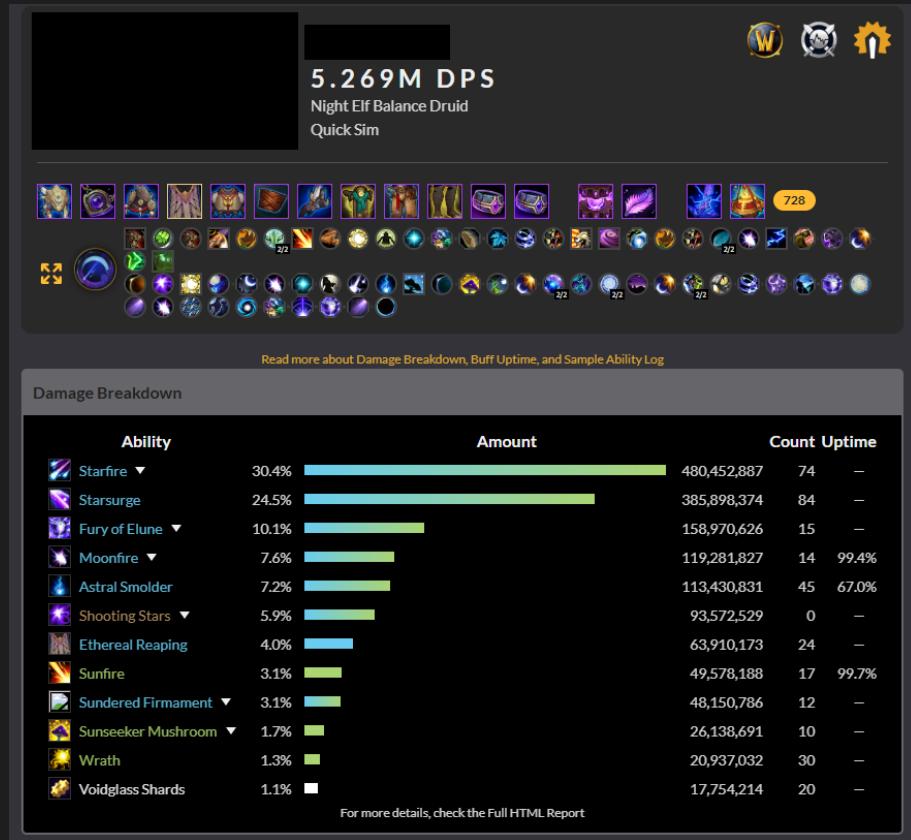
### Buff

### Uptime

### Count

	Eclipse (Lunar)	91.1%		19
	Starlord	90.2%		123
	Gathering Moonlight	69.5%		39
	Balance of All Things (Arcane)	60.8%		19
	Parting Skies	60.2%		12
	Moonlight Suffusion	52.0%		13
	Fury of Elune	42.9%		24
	Umbral Inspiration	40.0%		23
	Solstice	37.2%		24
	Astral Antenna	35.5%		13
	Sundered Firmament	30.5%		12
	Araz's Ritual Forge	28.4%		3
	Umbral Embrace	28.1%		26
	Dreamstate	26.6%		18
	Keen Prowess	26.3%		7
	Flask of Alchemical Chaos (Mastery)	25.1%		3
	Flask of Alchemical Chaos (Haste)	25.0%		3
	Flask of Alchemical Chaos (Crit)	25.0%		3
	Flask of Alchemical Chaos (Vers)	24.9%		3
	Ascendance (Haste)	24.4%		9
	Ascendance (Vers)	24.3%		9
	Ascension (Crit)	24.3%		9
	Ascension (Mastery)	24.3%		9
	Astral Antenna (Orb)	23.5%		14
	Incarnation: Chosen of Elune	23.3%		4
	Eclipse (Solar)	23.3%		4
	Lunar Amplification	14.8%		55
	Balance of All Things (Nature)	14.7%		4
	Bloodlust	13.5%		1
	Tempered Potion	13.2%		1
	Warrior of Elune	11.1%		5

# Rule-Based Agent Simulation



# **Case Study: ML Agents for Testing Level Design**

---

# Using Agent-Based Models to Test Games

---

- **Goal:** Test level design hypotheses (spatial balancing)
  - Use [Unity's ML-Agents package](#)
  - Examples: Areas that should be easier to defend, areas that provide opportunity to infiltrate. I.e. the defending teams should have an advantage in their half of the map.
  - Example Hypothesis: A higher position allows the player to defend easier
- **Problem:** Gathering data to test hypotheses is difficult (many matches required, skill differences, etc.)
- Useful for competitive games (e.g., Counter-Strike, Valorant, League of Legends, Overwatch, ...)

# Using Agent-Based Models to Test Games

---



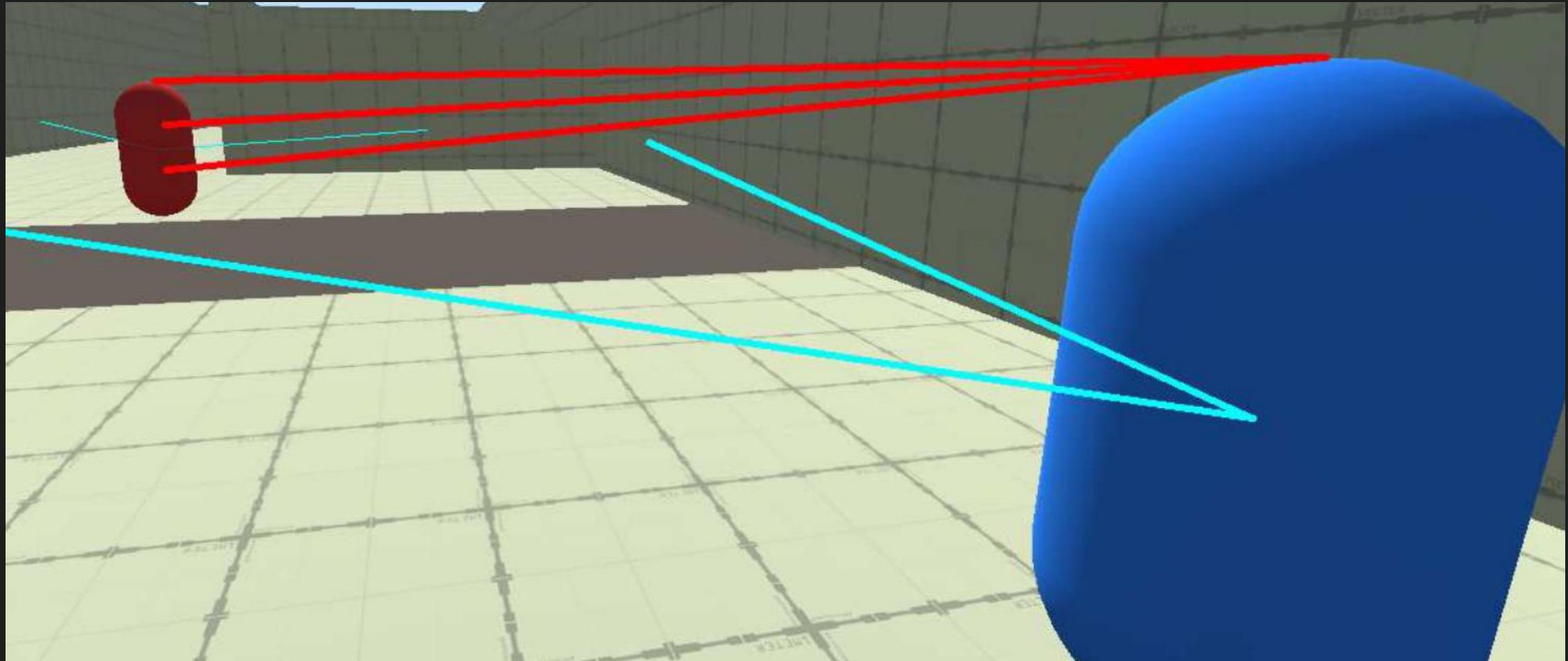
# Using Agent-Based Models to Test Games

---



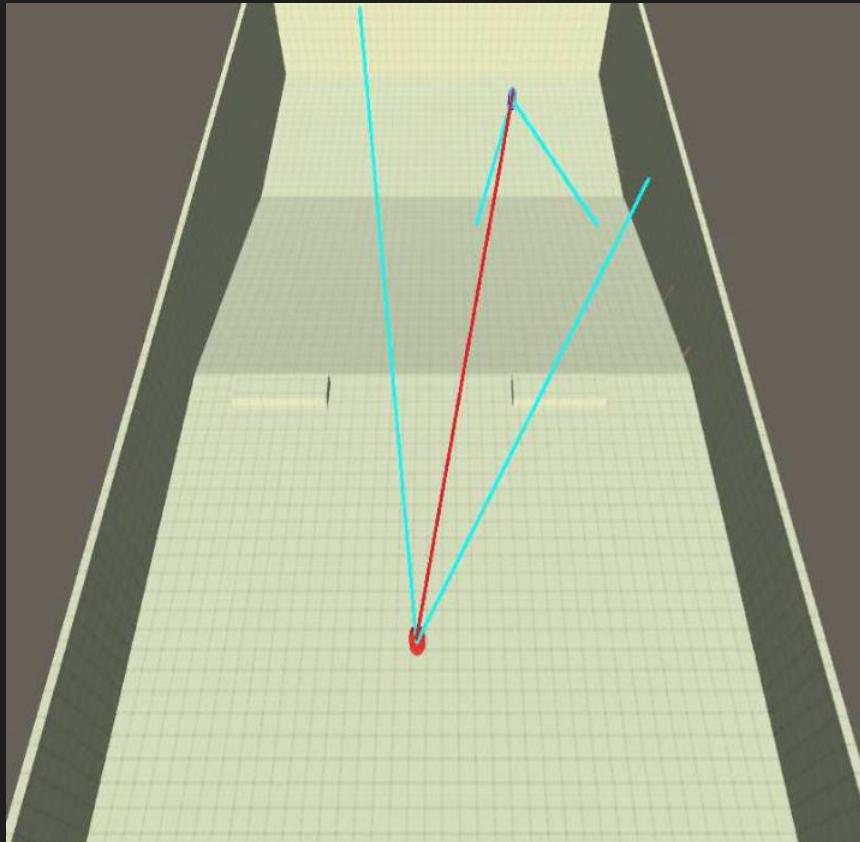
# Using Agent-Based Models to Test Games

---



# Using Agent-Based Models to Test Games

---



# Using Agent-Based Models to Test Games

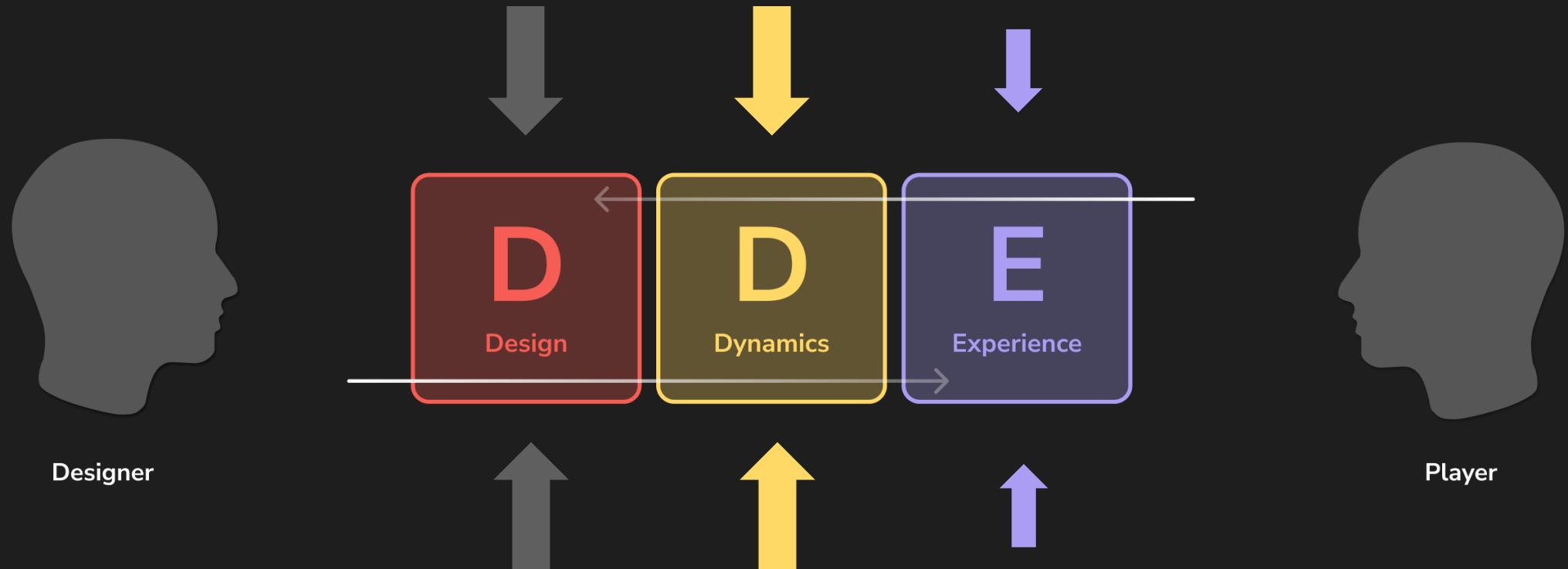


Source: Counter-Strike Series

2025 - 76 •

# Play-Testing – AI Approaches

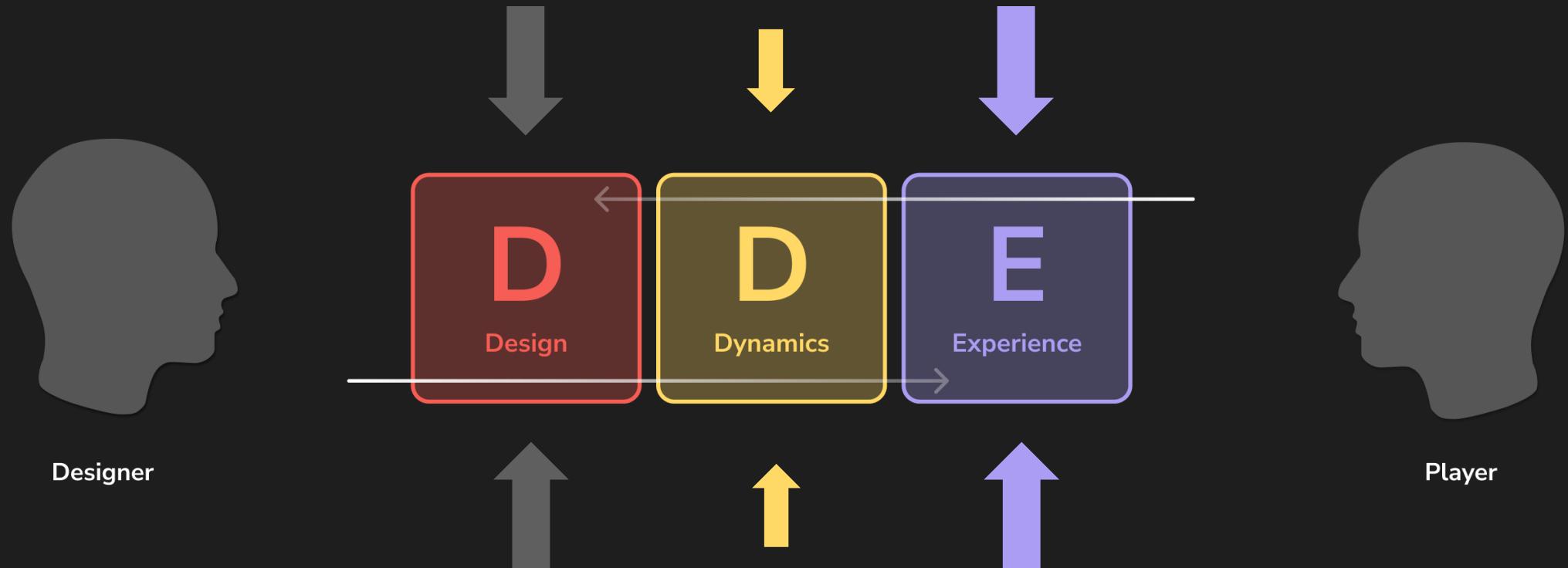
---



- “Expert” Feedback via LLMs
- AI playing games (finding bugs)
- AI playing games
- Simulating games
- AI processing and interpreting play data
- Playing with AI agents

# Play-Testing – AI Approaches

---



- “Expert” Feedback via LLMs
- AI playing games (finding bugs)

- AI playing games
- Simulating games

- AI processing and interpreting play data
- Playing with AI agents

# **Case Study: Evaluating the (Measured) Experience**

---

# Case Study: Evaluating the (Measured) Experience

---

**General Idea:** log game data (and more) and then analyze the data to draw conclusions about the dynamics and possible experience of players.

- System data
- Behavioral data
- Psychophysiological data

- Case Study: World of Warcraft Analyzer Tool (+ extend this approach with AI approaches)
- Case Study: XIVAnalysis for Final Fantasy XIV (+ extend this approach with AI approaches)

# **Case Study: World of Warcraft Analyzer Tool**

---

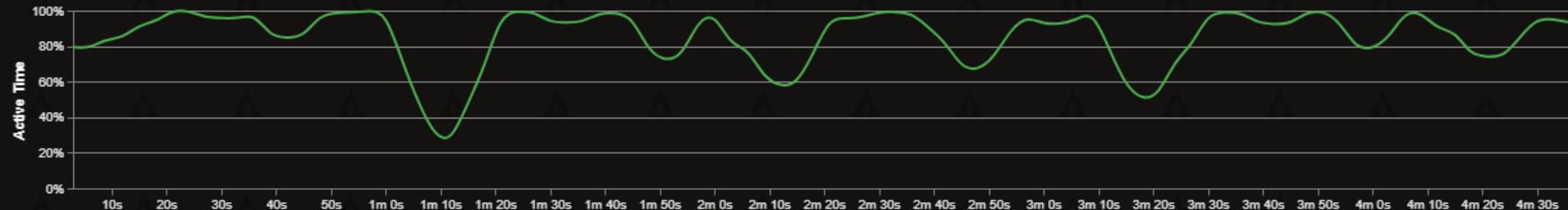
# Using Rule-Based Models to Provide the Player with Feedback (WoW)

## Always be Casting

***Continuously chaining casts throughout an encounter is the single most important thing for achieving good DPS as a caster.***

There should be no delay at all between your spell casts, it's better to start casting the wrong spell than to think for a few seconds and then cast the right spell. You should be able to handle a fight's mechanics with the minimum possible interruption to your casting. Some fights have unavoidable downtime due to phase transitions and the like, so in these cases 0% downtime will not be possible - do the best you can.

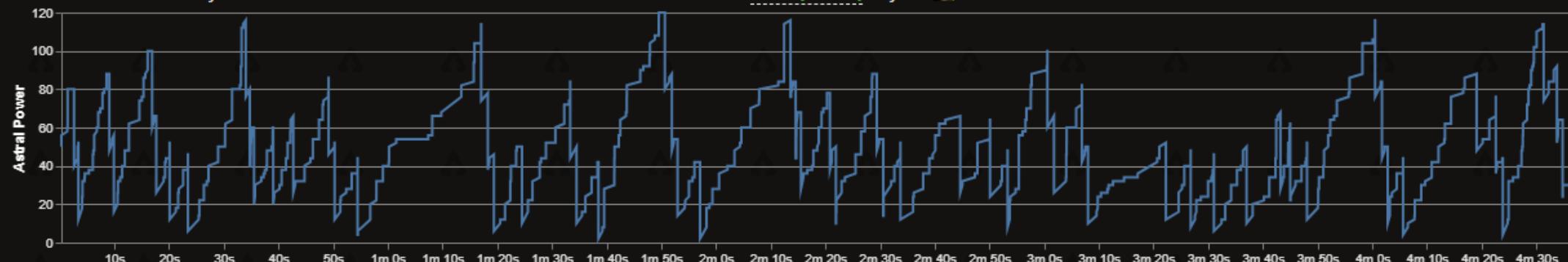
Active Time: 85.3% ✘ Cancelled Casts: 12.2% ✘



## Astral Power

Your primary resource is 🟤 Astral Power. Most of your spells generate Astral Power, which can be spent to cast ⚡ Starsurge or ⚡ Starfall. Avoid capping Astral Power!

The chart below shows your Astral Power over the course of the encounter. You wasted ✓ 14 (0.72%) of your 🟤 Astral Power.



## Rotation

Balance's core rotation involves maximizing time spent in **Eclipse**, maximizing DoT uptimes, and spending Astral Power to avoid overcapping. After fulfilling these priorities, open GCDs are filled with **Wrath** or **Starfire** depending on Eclipse type and target count. Refer to the spec guide for more [rotation details](#).

✿ **Moonfire** and ✿ **Sunfire** are high damage-per-cast-time DoTs that further boost your spell damage via Mastery. Maintaining 100% uptime is your highest priority.

### DoT Uptimes



▣ **Eclipse** dramatically increases your damage while active - when it fades you should immediately reactivate it.

It's best to enter ▲ **Eclipse (Lunar)** when you can hit 3 or more stacked targets with ✿ **Starfire**, and best to enter □ **Eclipse (Solar)** otherwise.

### Eclipse uptimes



Filler spells are ✿ **Wrath** and ✿ **Starfire**.

They are spammable and generate Astral Power. Use ✿ **Wrath** in single target and ✿ **Starfire** against multiple stacked targets.

Your fillers are greatly buffed by their corresponding □ **Eclipse** - aim to enter an Eclipse that matches your current target count.

If you make a mistake and find yourself in Lunar Eclipse with no stacked targets or in Solar Eclipse with stacked targets, you should use ✿ **Wrath**.

### Filler cast breakdown

Green is a good cast, Yellow is a Wrath during Lunar Eclipse, Red is a bad Starfire. Mouseover for more details.

Spender spells are ✿ **Starsurge** and ✿ **Starfall**.

They spend Astral Power to do big damage. Use ✿ **Starsurge** against 1 target, and ✿ **Starfall** against multiple targets.

Never use spenders outside of □ **Eclipse**.

### Starsurge cast breakdown

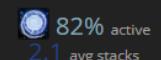
Green is a good cast, Red is without Eclipse active. Mouseover for more details.

### Starfall cast breakdown

Green is a good cast, Yellow hit too few targets, Red is without Eclipse active. Mouseover for more details.

No Starfalls cast this encounter!

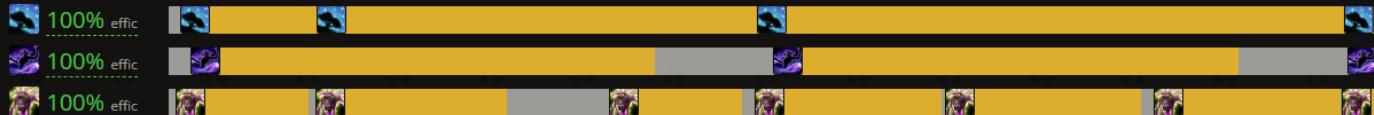
### Starlord uptime



## Cooldowns

Balance's cooldowns are moderately powerful and as with most DPS specs they should not be held for long. In order to maximize usages over the course of an encounter, aim to send the cooldown as soon as it becomes available (as long as you can be active on target over its duration).

**Cooldown Graph** - this graph shows when you used your cooldowns and how long you waited to use them again. Grey segments show when the spell was available, yellow segments show when the spell was cooling down. Red segments highlight times when you could have fit a whole extra use of the cooldown.



**Celestial Alignment** is our primary damage cooldown. It's best used as soon as it's available, but can be held to ensure you'll have full target uptime during its duration (don't use it when it will be interrupted by a fight mechanic).

**Per-Cast Breakdown** - click to expand

- @ 0:02 — **Celestial Alignment** — \*
- @ 0:33 — **Celestial Alignment** — \*
- @ 2:12 — **Celestial Alignment** — ⚡
- @ 4:25 — **Celestial Alignment** — ✓

## Preparation

### Enchants

Enchantments are easy ways to improve your throughput.



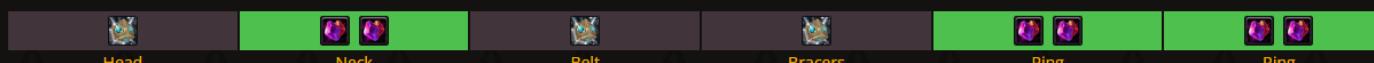
### Enhancements

Enhancements are easy ways to improve your throughput.



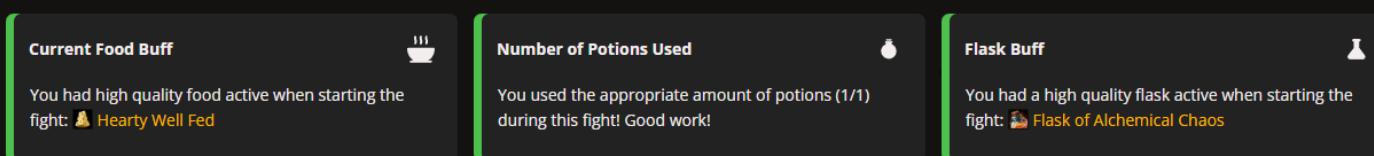
### Gems

Gems can increase a variety of stats. This indicates gear where you are missing Gem Sockets, don't have the highest crafted gems, or have empty sockets.



### Consumables

Using consumables appropriately is an easy way to improve your throughput.



## Defensives

Hide Explanations

Effectively using your major defensive cooldowns is an important aspect of your performance, as it will not only increase your own survivability, but also your entire raid by allowing healers to focus on keeping others alive.

As an Mage you have access to many defensives CDs such as Mirror Image, Greater Invisibility and Ice Block/ Ice Cold.

There are two things you should look for in your cooldown usage:

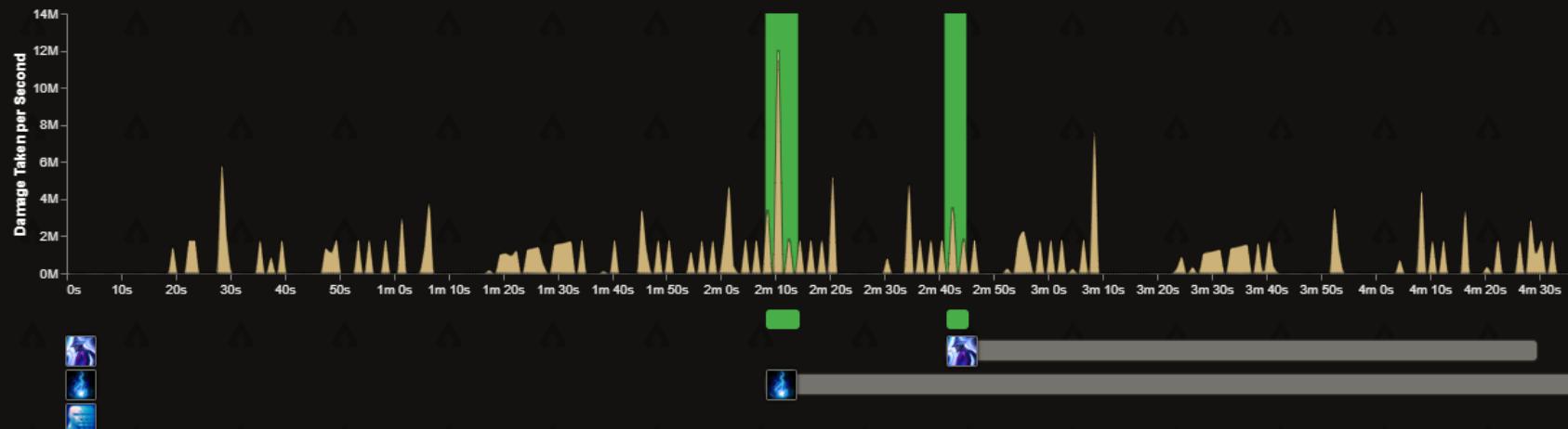
1. You should cover as many damage spikes as possible, and use any left over to cover periods of heavy, consistent damage.

In the damage chart below, a spike highlighted in green was covered by a defensive.

2. You should *use* your cooldowns. This may seem silly—but not using defensives is a common problem! For Mages, it is also likely to be fatal, since most of your mitigation lies in your active cooldowns.

Below the damage chart, your cooldowns are shown. Large gaps may indicate that you could get more uses—but remember that covering spikes is more important than maximizing total casts!

### Damage Taken



Mirror Image reduces the damage you take by 20%.

**Cast Breakdown** - These boxes represent each cast, colored by how much damage was mitigated. Missed casts are also shown in yellow or red.

Click on a box in the cast breakdown to view details.

Greater Invisibility reduces the damage you take by 60%.

**Cast Breakdown** - These boxes represent each cast, colored by how much damage was mitigated. Missed casts are also shown in yellow or red.

Click on a box in the cast breakdown to view details.

Ice Cold reduces the damage you take by 70%.

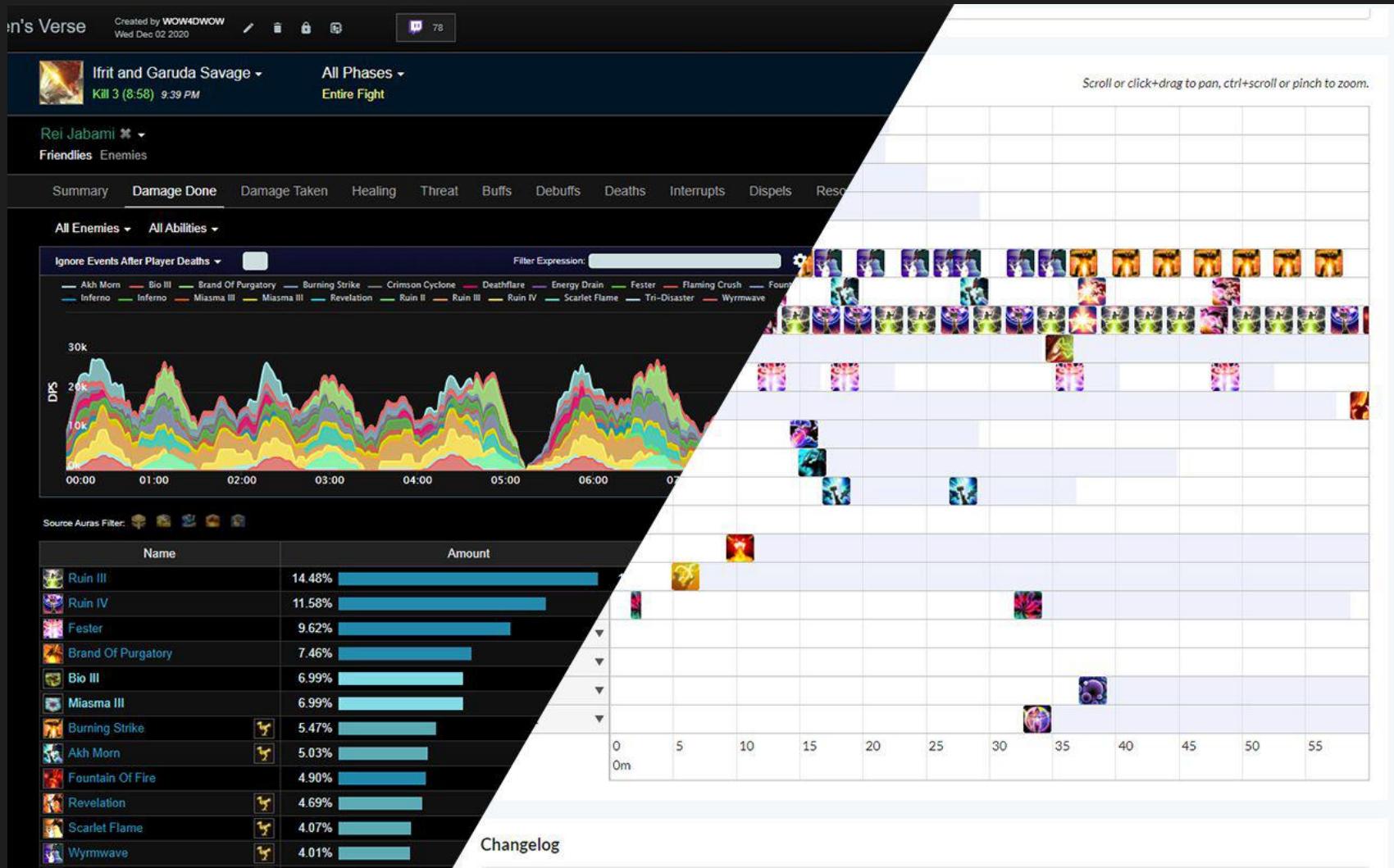
**Cast Breakdown** - These boxes represent each cast, colored by how much damage was mitigated. Missed casts are also shown in yellow or red.

Click on a box in the cast breakdown to view details.

# **Case Study: Analysis for Final Fantasy XIV**

---

# Using Rule-Based Models to Provide the Player with Feedback (FF)



## Checklist

### ✗ **Always be casting**

95.4%

**i** Make sure you're always doing something. It's often better to make small mistakes while keeping the GCD rolling than it is to perform the correct rotation slowly.

- GCD Uptime: 95.42%

### ▶ ✓ Use your cooldowns

98.7%

### ▶ ✓ Keep your Standard Finish buff up

100.0%

### ▶ ✓ Choose a Dance Partner

100.0%

## Suggestions



Show minor



Avoid breaking combos, as failing to complete combos costs you a significant amount of DPS and important secondary effects.

Using a combo GCD at the wrong combo step, using non-combo GCDs while inside a combo, missing, or attacking a target that is invulnerable will cause your combo to break.

Medium

You misused 2 combo actions.



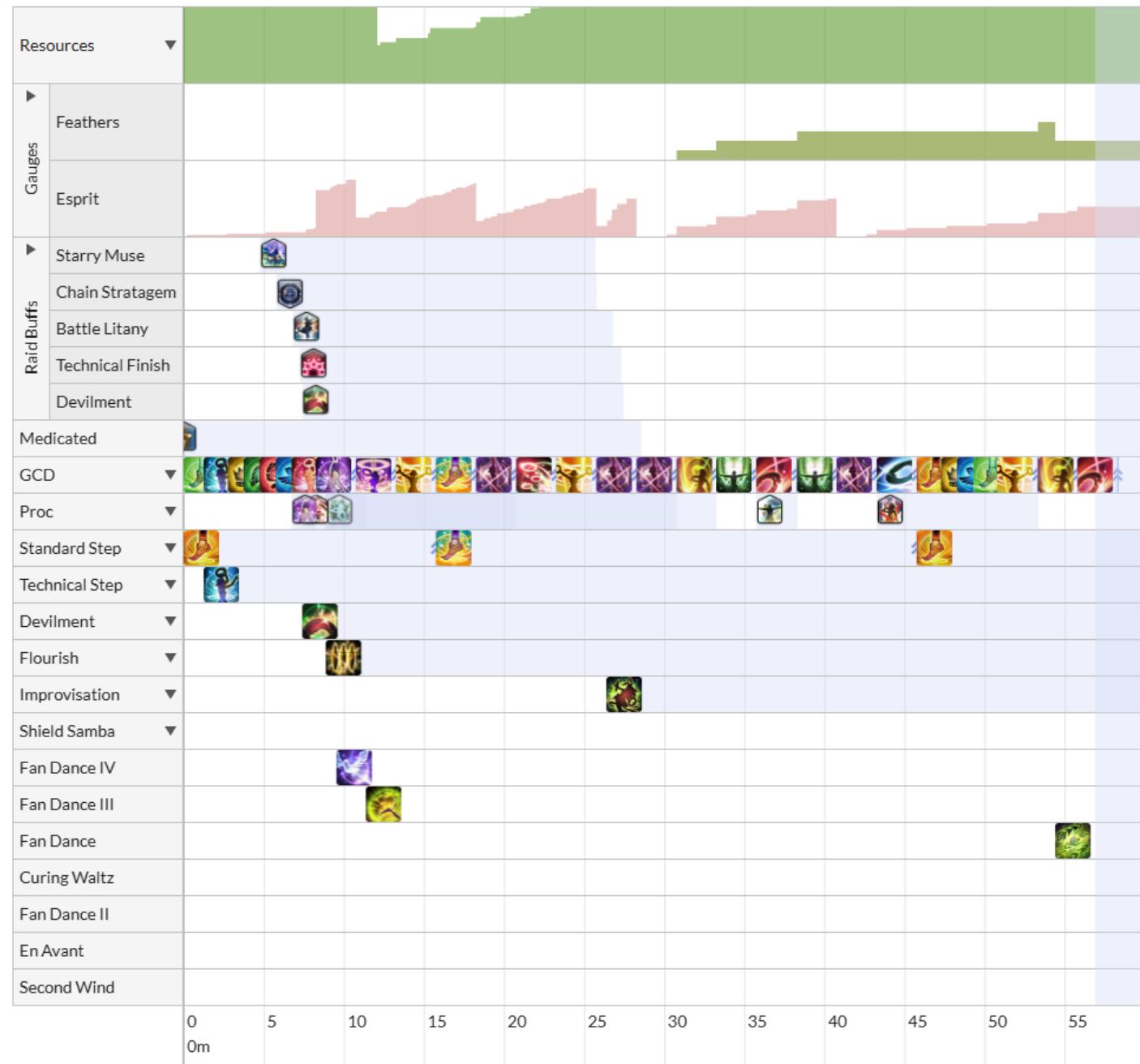
Estimated GCD

2.50s



## Timeline

Scroll or click+drag to pan, ctrl+scroll or pinch to zoom.



## Technical Windows

Time	Players Buffed	GCDs							Rotation		On Time?		Pooled?
00:07 	8/8	8/8	1/1	1/1	1/1	1/1	4/4	1/1					Opener
02:07 	8/8	8/8	1/1	1/1	1/1	1/1	4/4	1/1					
04:20 	8/8	8/8	1/1	1/1	1/1	1/1	3/3	1/1					
06:20 	8/8	8/8	1/1	1/1	1/1	1/1	4/4	1/1					
08:20 	8/8	8/8	1/1	1/1	1/1	1/1	3/3	1/1					

## Tinctures

Time	Rotation
< 0s	
05:39	

## Proc Issues

Time	Proc Status	Issue
04:42	Silken Symmetry	Dropped

## Combo Issues

Start Time	Combo Actions	Broken Time	Combo Breaker	Reason
04:12		04:42		Expired
05:20		05:46		Broken Combo

## Defensives



Using your mitigation and healing cooldowns can help you survive mistakes, or relieve some stress on the healers and let them deal more damage. While you shouldn't use them at the expense of your rotation or buff alignment, you should try to find helpful times to use them.

The below tables will show you where you can gain additional uses of these cooldowns, without interfering with your existing ones.

▶ Shield Samba - 5 / 5

▶ Improvisation - 4 / 4

▶ Curing Waltz - 3 / 8

▶ Second Wind - 2 / 4

# **Case Study: Analyzing Toxicity in Twitch Live Stream Chats**

---

# Analyzing Toxicity in Twitch Live Stream Chats

---

**Idea:** Learning from Twitch Live Streams for Player Experience

Observing people playing a game and observing people watching other people playing a game might provide valuable information for us!

# Analyzing Toxicity in Twitch Live Stream Chats

---

- Understanding and researching triggers and patterns behind toxic behavior on the Internet is crucial given the significant growth in online interactions worldwide
- Toxic behavior, manifesting as aggression, harassment, and bullying [6], is pervasive across various online platforms, including social media, comment sections, blogs, in-game chats, and live stream chats [4], [5] with an increased presence of trolls, bullies, and toxic individuals
- Problems
  - i) platforms struggle to recognize or delete toxic content
  - ii) struggle to identify **persistent** toxic users
- Potential solution in live stream chats: automatic detection and deletion of abusive language or by employing moderators who manually remove inappropriate messages and enforce chat rules
- Preventive measures like chatbots and content filters

# Analyzing Toxicity in Twitch Live Stream Chats – Methodology

---

- Developed a Python script to connect Twitch IRC interface using sockets. Results in a **dataset comprising chat messages posted during a stream.**
- Logged approximately 100,000 messages, averaging around 3,000 messages per streamer. 36 Twitch channels with typically 4-5 hours of streaming per streamer
- Selection Criteria: channel size (number of followers), gender, streamed content
- Data Pre-Processing:
  - i) organized structured data frame with essential attributes, including timestamp, username, message content
  - ii) identifying type of each message entry by introducing categorization attribute

TABLE I  
SIZE CATEGORIES OF TWITCH CHANNELS

Category Label	Number of Followers
SMALL	$\leq 5,000$
AVG	$\leq 25,000$
BIG	$\leq 1,000,000$
XXL	$> 1,000,000$

# Analyzing Toxicity in Twitch Live Stream Chats – Methodology

---

- Natural Language Processing (**NLP**) Techniques for Data Analysis
- **Sentiment Analysis with VADER model** to determine whether the emotion conveyed is positive, negative, or neutral [11]. VADER provides a polarity score, representing the ratios of positive, negative, and neutral sentiments, which sum up to one. Notably, VADER has shown comparable, or even superior performance compared to other highly regarded sentiment analysis tools [12]
- To identify which messages are part of toxic communication, we made use of the **open-source classifier model Detoxify** [13], which is trained especially for toxic comment classification.
- The collected data underwent a normality check using the Shapiro-Wilk test. Subsequently, depending on the normality result, significance testing was performed using either one-way ANOVA (for normally distributed data) or Kruskal-Wallis (for non-normally distributed data). Both tests employed a significance threshold of 0.05 to evaluate the null hypothesis that all means are equal.

## Analyzing Toxicity in Twitch Live Stream Chats – Findings

---

**Summary of findings:** "In this study, we provide a dataset of chat logs from various streamers and their chat activities. Our initial findings indicate that gender, community size, and game genre influence the occurrence of toxic behavior in the chat. Male streamers and larger communities tend to have more toxic interactions. Additionally, multiplayer and shooter games exhibit a higher frequency of toxic messages compared to single-player games, strategy games, or role-playing games. We also identify potential triggers, including text-based and in-stream factors."

# **Case Study: Substituting Players**

---

## **Substituting Players – Development of Battlefield V Bots**

---

- Multiplayer Client Stability Testing
- AI delivers input based on game context
- AI agents have different objectives and tasks (e.g., defend aggressively vs. defensively)
- AI can be used to fill in for actual players
  
- More information in the GDC Talk

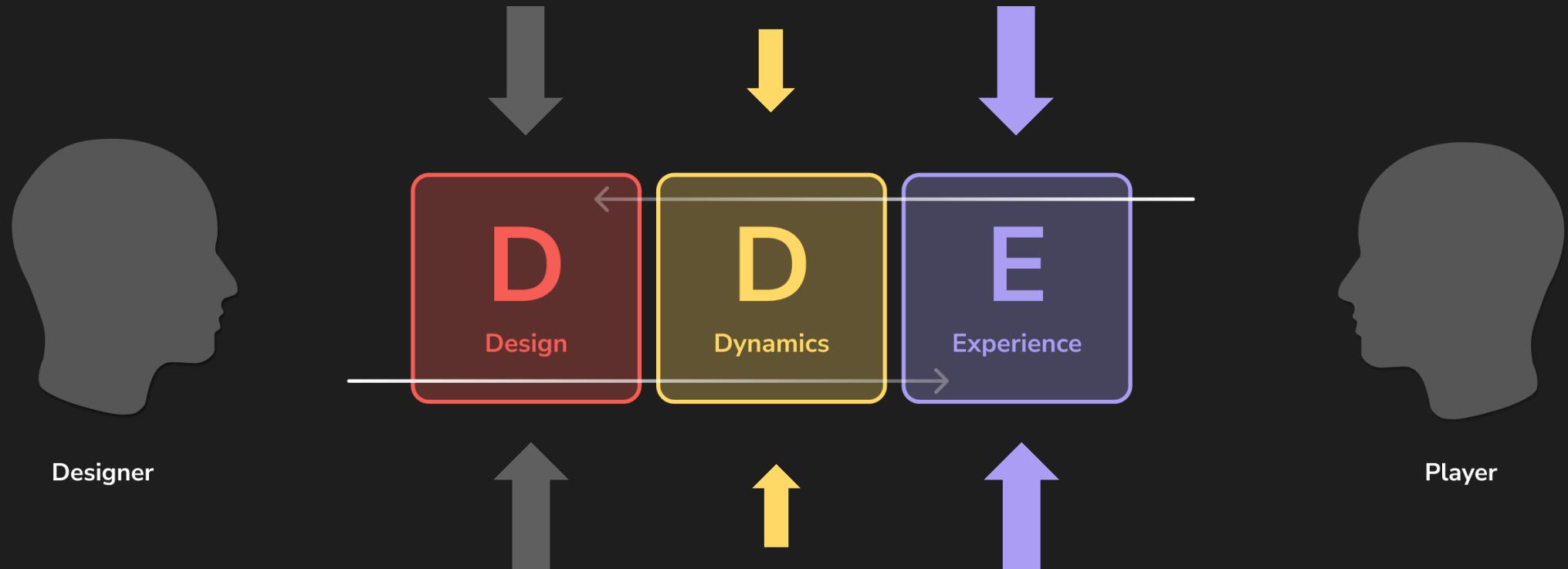
## Substituting Players – Other Notable Examples

---

- **Dead by Daylight** – When a living survivor disconnects, a bot takes over for that player
- **Rocket League** – Disconnected player gets taken over by AI (even in ranked mode)
- **Counter Strike** (non-ranked) – A bot takes over for a disconnected player; limited ability to control the bot with voice commands
- **Super Smash Bros. Ultimate** – AI can substitute player's for both normal play as well as for the training mode

# Play-Testing – AI Approaches

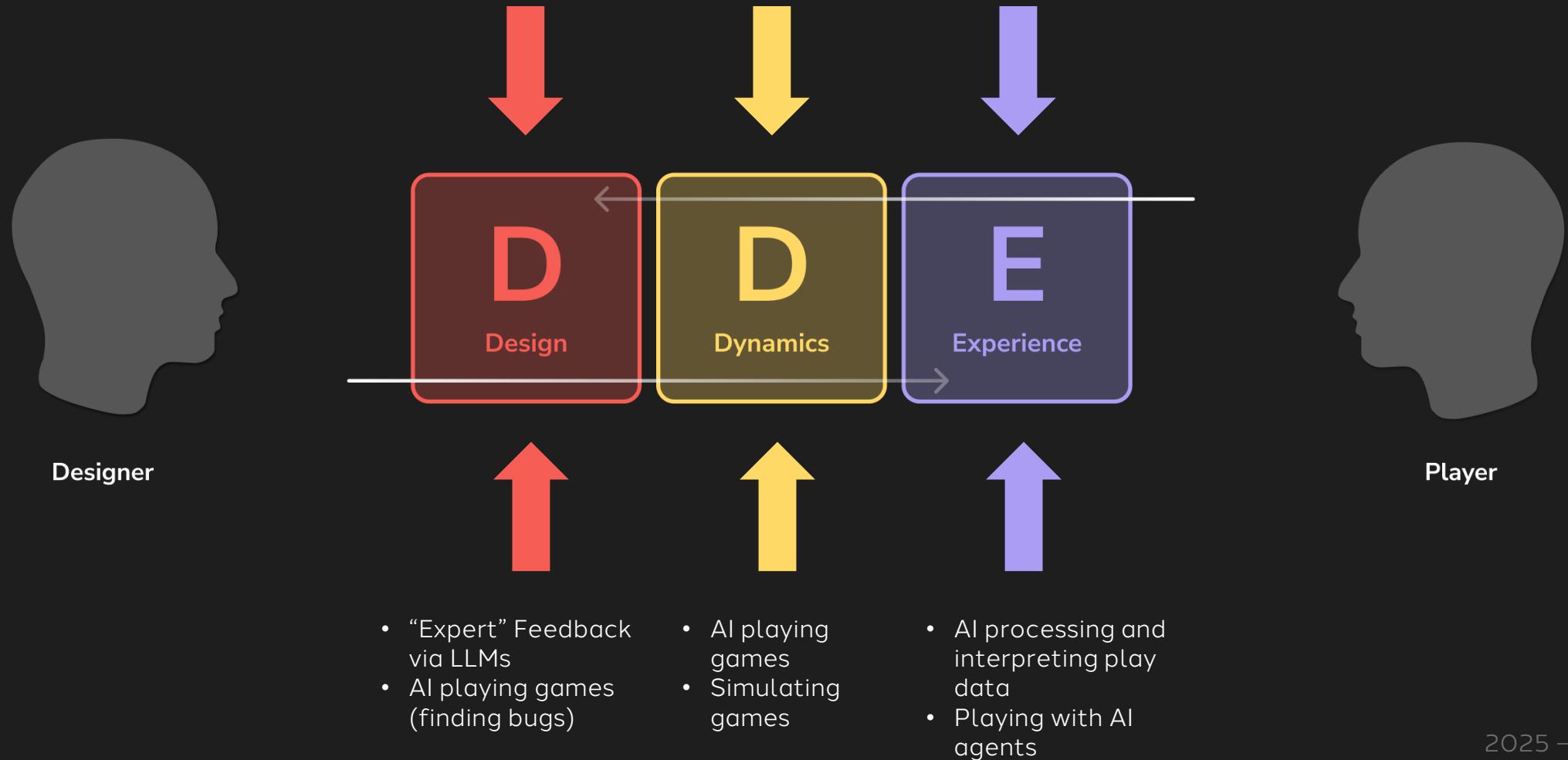
---



- “Expert” Feedback via LLMs
- AI playing games (finding bugs)
- AI playing games
- Simulating games
- AI processing and interpreting play data
- Playing with AI agents

# Play-Testing – AI Approaches

---



## Other Approaches

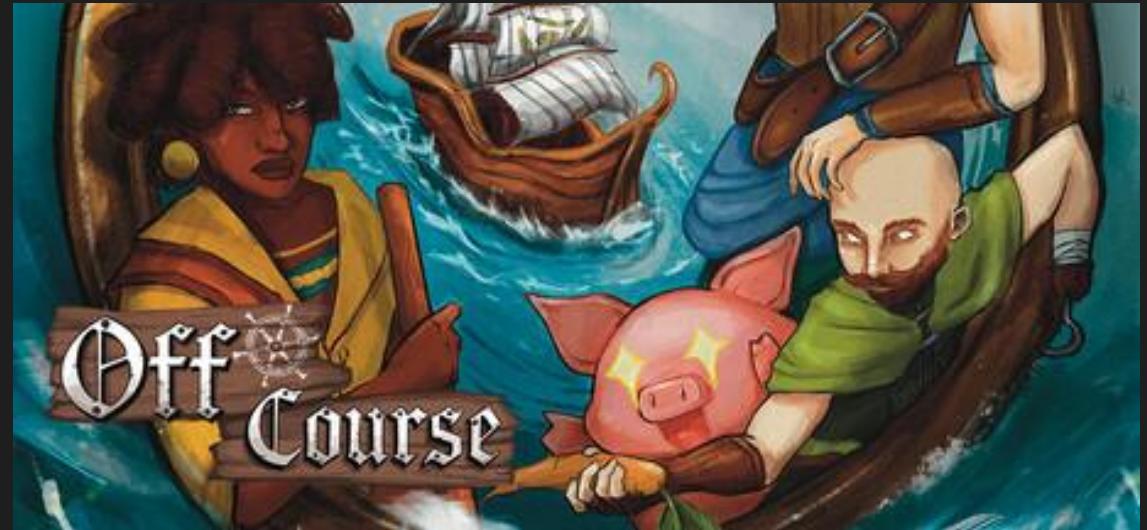
---

- Limited Case Studies by Daniel
- Of course there is more!

## Other Approaches

---

- Limited Case Studies by Daniel
- Of(f) course there is more!



# Further Reading

---

## Further Reading (Watching)

---



[Valve's "Secret Weapon" – Game Maker's Toolkit \(YouTube\)](#)



[Brackeys "How to playtest"](#)

## Further Reading

---

For more information about user studies, we refer to the course “**3D User Interfaces**” and the book “**Research Methods in Human-Computer Interaction**”.

### RESEARCH METHODS IN HUMAN-COMPUTER INTERACTION

Second Edition



Jonathan Lazar | Jinjuan Heidi Feng | Harry Hochheiser

# References

---

## References

---

[Valve's "Secret Weapon", Game Maker's Toolkit](#)

[Artificial Games and Intelligence, Yannakakis, Georgios N. and Togelius, Julian](#)

[Toward a Game Design Engineering Process Centered on Player Experience, Dyrda, Daniel and Klinker, Gudrun](#)

[Diamonds in the rough: Transforming SPARCs of imagination into a game concept by leveraging medium sized LLMs, Geheeb, Julian and Ivan, Farhan Abid and Dyrda, Daniel and Anschütz, Miriam and Groh, Georg](#)

[Playtesting Without Humans: Agent-Based Modeling as Testing Tool for Game Spaces, Neske, Marvin and Dyrda, Daniel](#)

# Thank You!

---