A RESEARCH PROPOSAL

Transparency in Emergent Narrative: Explainable AI Models for GOAP Agents and AI Storytellers

Süleyman Ceran
Mobile: +90 542 115 33 74
E-mail: **suleymansceran@gmail.com**
Institution: TOBB UNIVERSITY OF ECONOMICS & TECHNOLOGY
Deparment: Computer Engineering, Graduate

Transparency in Emergent Narrative: Explainable AI Models for GOAP Agents and AI Storytellers

1. Introduction

1.1. The Rise of Emergent Narrative in Interactive Entertainment

Emergent Narrative (EN) represents a paradigm shift in interactive storytelling, where narratives are not explicitly pre-authored by developers but arise dynamically from the player's interactions with complex game systems and Artificial Intelligence (AI)-driven entities.[1] This approach contrasts sharply with traditional linear or branching narratives by offering unique, deeply personal, and player-driven story experiences.[1] The significance of EN is increasingly recognized in modern game design, evidenced by its successful implementation in seminal titles such as *The Sims* [2], *Crusader Kings* [5], *Rimworld* [8], and *Dwarf Fortress*.[7]These games compellingly demonstrate EN's capacity to create highly replayable and personally meaningful experiences, wherein the "story is constructed by the player, through his (inter)actions and explorations".[1] The principles articulated by Fahraeus, such as the critical roles of sophisticated AI actors and inherent unpredictability in fostering EN [5], underscore the systemic foundations of these narratives.

The fundamental appeal of EN stems from its ability to generate what Fahraeus terms "game anecdotes" [5] or what others describe as "accidental gameplay".[3] These are unique sequences of events and experiences over which players feel a strong sense of ownership, effectively transforming them from passive consumers of a pre-defined story into active co-creators of their own unfolding sagas.[2] Games like *Minecraft*, for example, illustrate that players often crave their own unique experiences and the opportunity to forge their own paths, rather than merely consuming developer-scripted content.[3] Emergent narrative systems, through their intricate systemic interactions [1], are uniquely positioned to facilitate this desire for player-authored experiences.[2] This capacity for fostering player agency and personal narrative construction is a core strength of EN and a primary driver of its growing prominence in contemporary game design.

1.2. The Role of Complex AI: GOAP Agents and AI Storytellers

The engines driving these rich emergent narratives are, invariably, complex Artificial Intelligence systems. This research proposal specifically focuses on two key architectural components that are instrumental in creating EN: Goal-Oriented Action Planning (GOAP) agents and AI Storytellers.

GOAP agents, famously demonstrated in games like Monolith Productions' *F.E.A.R.* [13], empower non-player characters (NPCs) with the ability to dynamically formulate and execute sequences of actions to achieve specified goals. This contrasts with traditional, more rigid AI scripting techniques, allowing for believable, adaptive, and often unpredictable NPC behaviors that are crucial for the dynamism of EN.[12] The capacity of GOAP agents to adapt their plans based on evolving world states and unforeseen circumstances makes them powerful tools for generating dynamic and responsive character actions within the game world.[16]

Complementing the agent-level intelligence of GOAP, AI Storytellers operate at a higher, more abstract level of narrative management. Exemplified by systems like the AI Director in Valve's *Left 4 Dead* [17] and the various storyteller personas in Ludeon Studios' *Rimworld* [8], these AI systems are responsible for managing the overarching narrative flow, pacing, and the introduction of significant events or challenges. Their function is to shape the player's experience, maintain a degree of narrative coherence, or inject desired dramatic tension, often acting as an "unseeable God" [8] that subtly or overtly influences the game world and its unfolding events.

The synergy between autonomous, goal-driven agents (GOAP) operating at a micro-level, generating moment-to-moment behaviors [12], and a guiding AI Storyteller operating at a macro-level, orchestrating larger narrative arcs and challenges [8], creates a fertile and rich ecosystem for emergent narratives to flourish. However, it is precisely this multi-layered complexity, with its intricate interplay of autonomous decisions and systemic interventions, that contributes significantly to the opacity of these systems, a challenge this research aims to address.

1.3. The "Black Box" Problem in Emergent Narrative Systems

Despite their power in generating compelling EN, these sophisticated AI systems—GOAP planners and AI Storyteller rule engines—often suffer from the "black box" problem: their internal decision-making processes are opaque to those who design and develop them.[21] This lack of transparency presents considerable challenges from a game development and design standpoint.

Firstly, **designing** for specific types of emergent experiences becomes exceedingly difficult if the underlying AI's reasoning remains obscure. Game designers may struggle to predict how modifications to rules, parameters, or agent capabilities will manifest in terms of AI behavior and, consequently, the emergent narrative.[22] Secondly, **debugging** undesirable AI behaviors—such as an agent consistently failing to achieve a critical goal, or an AI Storyteller generating an unfairly punitive or nonsensical sequence of events—becomes an arduous task.[12] As noted by Duffy in the context of developing EN systems, debugging GOAP agents was "especially time consuming" due to the multiplicity of potential outcomes from identical starting conditions.[12] Thirdly, **balancing** the game's difficulty, pacing, and the overall emergent experience is a delicate act made more challenging when the AI's internal state, motivations, and decision-making criteria are not visible to the development team.[8]

While not the primary focus of this Master's project, this opacity can also extend to the **player experience**. Players might find it difficult to understand why certain events unfold as they do. If the "magic" of emergence devolves into perceived random chaos or arbitrary system behavior, it can lead to player frustration or a diminished sense of agency and narrative coherence.[19]

The "black box" nature of these AI systems is therefore not merely a technical inconvenience; it represents a fundamental barrier to the maturation and intentional design of rich, robust, and reliable emergent narrative systems. If designers and developers cannot effectively peer into the workings of these AI components, they are, in essence, "designing in the dark," relying more on laborious trial-and-error methodologies rather than principled, iterative refinement. The desire for richer EN pushes towards more complex AI [1], which in turn inherently leads to greater opacity.[21] This opacity then directly hinders the crucial development processes of iterative design, debugging, and balancing.[12] Consequently, to advance the design of EN systems beyond current practices, methods to reduce this opacity and enhance transparency are essential.

1.4. Explainable AI (XAI) as a Potential Solution

Explainable AI (XAI) is an evolving field within Artificial Intelligence dedicated to developing techniques and models that render AI decision-making processes transparent and understandable to human users.[25] It is posited that XAI principles and techniques—such as model-specific explanations, visualization tools, rule extraction, and the development of interpretable-by-design models—can be effectively applied to address the "black box" problem inherent in the GOAP agents and AI Storytellers that drive emergent narratives.[27]

The primary focus of applying XAI in this research context is to empower game designers and developers. By providing them with tools and insights into the internal workings of these complex AI systems, XAI can facilitate more informed design decisions, more efficient debugging processes, and more effective balancing of the emergent experience. It is important to note that the application of XAI here is not intended to "dumb down" the AI or to entirely remove the "magic" of emergence for the player. The goal of EN often involves player-driven discovery, surprise, and the joy of interpreting unforeseen events.[1] Exposing all AI internals directly to the player could indeed "spoil the magic" and undermine the sense of wonder.[24] Instead, XAI's role in this project is conceived primarily as a suite of developer-facing tools and methodologies, aimed at providing design-time and debug-time clarity for the creators of these EN systems.

1.5. Core Aim and Research Contribution

The core aim of this Master's research project is: *To explore how Explainable AI (XAI) principles and techniques can be practically designed and implemented for Goal-Oriented Action Planning (GOAP) agents and AI Storytellers to enhance transparency from a game design and development perspective, focusing on the creation of prototype XAI tools and the derivation of design guidelines.*

The primary contribution of this research will be to provide valuable design insights and practical, prototyped examples of XAI application within these specific game AI architectures. This work seeks to address an identified gap in the current research landscape by focusing on XAI solutions tailored to the unique challenges of developing and understanding AI systems that generate emergent narratives.

2. Background and Literature Review

2.1. Emergent Narrative in Games

Emergent Narrative (EN) arises from the dynamic interplay of game mechanics, AI-driven behaviors, and player actions, rather than being explicitly pre-scripted.[1] In games like *The Sims*, players effectively "write their own stories" through their interactions with autonomous characters and complex underlying systems.[2] This player-centric construction of narrative is a hallmark of EN. The concept also extends to "spatial storytelling," as seen in games like *Minecraft*, where players forge narratives through exploration, interaction, and modification of a procedurally generated world, constructing meaning from their unique journey and experiences within the game space.[2] The significance of EN lies in its capacity for deep personalization and replayability, as each playthrough can yield a distinct narrative tapestry woven by the player's choices and the system's responses.

Several key theories and principles underpin the design of EN systems. Henrik Fahraeus, the designer of *Crusader Kings II*, has articulated several ingredients for compelling emergent drama. These include sandbox gameplay offering significant player freedom, a multitude of AI actors possessing distinct personalities and opinions, dynamically changing game conditions, inherent conflict, and an environment where "low morals" can lead to interesting transgressions and consequences.[5] Fahraeus emphasizes that "emergent stories allow for enormous replayability" and that "plausible AI actors are crucial" for the believability and richness of these narratives.[5] A critical insight from his work is that "the 'art' is expressed through the simulation itself, not any single emergent story" [5], highlighting the systemic and processual nature of EN rather than a focus on discrete, pre-authored plot points.

Henry Jenkins' work distinguishes emergent narratives, where players actively co-construct the story, from embedded narratives, which are pre-authored by designers.[2] This distinction is particularly relevant when analyzing games like *The Sims*. Furthermore, Jenkins and others have noted the potential for powerful experiences when both emergent and embedded narrative elements are thoughtfully combined within a single game.[2]

The player's role in EN is not passive; it is one of active interpretation and meaning-making.[1] As one source notes, "Without the act of collecting these events and constructing a narrative theme, the story events are hollow".[1] This implies that the game systems generate a stream of events, but it is often the player who acts as the "author," linking these events into a coherent and personally significant story. This cognitive effort by the player is a defining characteristic of the EN experience.

Revisiting examples, *Crusader Kings II* showcases EN through complex character interactions, where AI-driven motivations related to traits, relationships, and ambitions can lead to unscripted events like assassinations, alliances, betrayals, and revolts, often with a strong element of chance influencing outcomes.[5]*Rimworld* employs AI Storytellers—Cassandra Classic, Phoebe Chillax, and Randy Random—each with a distinct personality, to influence game events based on a multitude of factors such as colony wealth, colonist population, and recent occurrences, thereby shaping the unique narrative of each colony's survival.[8] *Dwarf Fortress* is renowned for its deeply intricate world generation and simulation, where complex interactions between creatures, environmental factors, and the player-managed dwarves lead to rich, often unpredictable, fortress-specific narratives.[7]

A central tension in designing for EN lies in balancing the systemic generation of potentially narrative-rich events with the player's cognitive capacity and motivation to weave these events into a coherent and engaging story. While the game systems provide the "story events" [1], the player often assumes the role of an author, actively linking and interpreting these occurrences.[1] If the generated events are perceived as too random or nonsensical, the player's ability to form a satisfying narrative may be compromised. Conversely, if events become too predictable, the sense of emergence and discovery is lost. Therefore, tools that help designers understand precisely how their AI systems (such as GOAP agents and AI Storytellers) generate these foundational events are crucial for striking this delicate balance. Explainable AI tools can offer designers visibility into the "raw material" of emergent narrative that their systems are producing, enabling more intentional design and tuning of the EN experience.

2.2. Goal-Oriented Action Planning (GOAP) in Games

Goal-Oriented Action Planning (GOAP) is an artificial intelligence planning technique increasingly utilized in game development to create more autonomous and believable non-player characters (NPCs). Unlike traditional Finite State Machines (FSMs) which rely on pre-scripted behaviors for every conceivable situation, GOAP allows agents to dynamically determine a sequence of actions to achieve a specific goal.[13] It represents a declarative approach to AI: "A planning system tells the A.I. what his goals and actions are, and lets the A.I. decide how to sequence actions to satisfy goals".[13] This empowers NPCs with a greater degree of flexibility and adaptability.

The core architecture of a GOAP system comprises several key components. **Actions** are fundamental units of behavior, each defined by a set of **preconditions** (conditions that must be true in the game world for the action to be executable) and **effects** (the changes to the world state that result from the action's execution).[13]Actions are also typically associated with a **cost**, representing the resources or effort required to perform them. **Goals** represent desired world states that an agent aims to achieve.[13] The **planner** is the reasoning component of GOAP, typically employing a search algorithm such as A* to find an optimal (e.g., lowest total cost) sequence of actions that transitions the agent from its current world state to a desired goal state.[13] The planner operates by matching the effects of available actions to the unmet preconditions of other actions or the goal itself, thereby constructing chains of actions that form a complete plan.[39]

Jeff Orkin's work on the AI for the game *F.E.A.R.* serves as a seminal example of GOAP's application in a commercial title.[13] The AI in *F.E.A.R.* utilized GOAP in conjunction with a remarkably simple FSM (consisting of only three states) and the A* algorithm for action planning. This architecture enabled NPCs to exhibit complex, adaptive, and tactically sophisticated squad-based behaviors, such as laying down suppression fire, strategically advancing on the player, and flanking maneuvers.[13] A key insight from the

*F.E.A.R.*implementation is that GOAP provides a robust framework for managing the inherent complexity that arises from combining numerous individual behaviors into a coherent and intelligent whole.[13]

GOAP offers several advantages over traditional FSMs, particularly in the context of emergent systems. It provides significantly greater behavioral **flexibility**, as agents are not locked into rigid state transitions. It offers better **scalability** for managing complex behaviors, as new actions and goals can be added modularly. This results in more **realistic** and less predictable NPC actions because plans are generated dynamically in response to the current, often changing, state of the game world.[13]

However, GOAP also presents challenges. The search space for potential plans can become very large, particularly with a high number of available actions or complex world states, which can impact real-time performance.[16] Furthermore, debugging GOAP-driven systems can be notably difficult due to their dynamic and emergent nature; understanding why an agent chose a particular plan, or why a plan failed, can be non-trivial.[12]

The true power of GOAP lies in its fundamental decoupling of goals from the specific sequences of actions required to achieve them. This allows for emergent solutions where agents can discover novel or unexpected ways to satisfy their objectives based on the current context. However, this very decoupling is what makes it challenging for designers to predict precisely which sequence of actions an agent will choose in a given situation, or to diagnose why a plan might be failing. This opacity in the planning process is a prime area where XAI interventions can provide significant value. GOAP agents essentially search a graph of possible actions to satisfy a goal.[13] The "optimal" path through this graph, as determined by an A* planner, is contingent upon action costs and the current world state.[16] Even minor alterations in world state variables or action costs can lead to vastly different plans being generated. Without tools to visualize this search space or the specific factors influencing path selection and pruning, designers face considerable difficulty in debugging agent behavior or fine-tuning its performance.[15] XAI techniques, such as interactive plan visualizers, can directly address this challenge by making the planner's reasoning process more transparent.

2.3. AI Storytellers in Games

AI Storytellers, also referred to as "AI Directors" or "Drama Managers," are sophisticated systems designed to manage the high-level aspects of game pacing, difficulty, and narrative event generation. Their primary purpose is to dynamically influence the player's experience, fostering a sense of an unfolding drama or a coherent narrative journey, rather than allowing events to be purely random or statically scripted.[8] These systems aim to weave a more engaging pattern of events than might arise from chance alone, often by reacting to player actions and the current game state.[3]

The implementation strategies for AI Storytellers vary, but several common approaches exist. Many storytellers operate on **rule-based systems**, where specific events are triggered or game parameters are adjusted when certain conditions are met. These conditions can include player progress, resource levels (e.g., colony wealth in *Rimworld*), the time elapsed since the last significant event, or metrics designed to estimate player stress or engagement levels.[8]

A key function of AI Storytellers is **pacing control**—the artful modulation of tension to create peaks and troughs in the player's experience. The AI Director in *Left 4 Dead* is a notable example, dynamically managing the population of common and special infected, as well as item spawns, based on factors such as player performance and their "flow distance" through the map.[17] The Director's goal is not simply to increase difficulty but to create dramatic tension, often described as a "tense mix of intense action and quiet anticipation".[19] It achieves this by tracking metrics like "emotional intensity" and transitioning through different internal states such as build-up, sustained peak, peak fade, and relax, each influencing its spawning behavior.[20]

AI Storytellers can also exhibit distinct **personalities or styles**, which dictate their event generation patterns and overall impact on the game's tone. *Rimworld* provides a clear illustration of this with its different storyteller options: Cassandra Classic, who creates a steadily increasing curve of challenge; Phoebe Chillax, who allows for more breathing room between major events; and Randy Random, whose unpredictable nature can lead to both highly fortunate and devastatingly unfair sequences of events.[8]

The impact of AI Storytellers on emergent narrative is significant. They act as a guiding hand, preventing the emergent story from becoming overly chaotic and nonsensical, or conversely, too dull and uneventful. By introducing "narrative seeds" in the form of challenges, opportunities, or unexpected events, they provide catalysts around which players can react and construct their own unique stories.[3]

AI Storytellers are, in themselves, complex decision-making systems. While their overarching goal is to shape the player experience in a desirable way, their internal logic—the rules they follow, the probabilities they consult, the state changes they undergo—can be opaque to game designers. This lack of transparency makes it difficult for designers to understand *why* a particular event was triggered at a specific moment, or *how* the storyteller is currently interpreting the game state (e.g., its current "emotional intensity" assessment in an L4D-like director, or the weighted probabilities for Randy Random's next dramatic event in *Rimworld*). Without this visibility, designers cannot effectively tune the storyteller's behavior to achieve the desired narrative pacing, nor can they easily debug situations where the pacing feels consistently off, or where the sequence of events seems unfair or illogical from a design perspective.[22] An XAI-driven dashboard for the AI Storyteller, revealing its internal state and decision-making parameters, could provide this crucial visibility, empowering designers to better understand and refine these powerful narrative-shaping systems.

2.4. Explainable AI (XAI) Techniques for Game AI

The core principles of Explainable AI (XAI) revolve around achieving transparency, interpretability, and the fundamental ability for humans to comprehend the decision-making processes of AI systems.[25] The overarching objective is to move beyond opaque "black box" models, where the internal workings are obscured, towards systems whose reasoning can be inspected and understood.[21] In the context of game AI, particularly for systems like

GOAP agents and AI Storytellers that drive emergent narratives, several XAI techniques hold promise for enhancing developer understanding and control.

A crucial category is **model-specific explainability**. These are techniques tailored to the particular architecture of the AI system in question. For GOAP agents, this could involve visualizations of the plan search space, highlighting the sequence of actions considered, the costs associated with different paths, and the reasons why certain plans were ultimately discarded in favor of the chosen one.[15] For rule-based AI Storytellers, model-specific explanations might involve displaying the currently active rules, the conditions that triggered them, and their individual contributions to the storyteller's decision to initiate an event or modify game parameters.[26]

Visualization stands out as a particularly powerful tool for XAI in game development.[27] This can encompass a range of approaches, from dashboards that display the real-time state of AI entities (e.g., current goals, mood, available resources for a GOAP agent; tension level, event probabilities for an AI Storyteller) [45], to dynamic graphs of internal parameters, or more abstract visual representations of plans, rules, or decision trees. Effective visualizations can make complex data streams and AI states significantly more accessible and interpretable for game designers.[44]

For more complex AI models, techniques like **rule extraction** or the creation of **surrogate models** can be beneficial.[26] Rule extraction aims to derive a set of simpler, human-readable rules that approximate the behavior of a more complex model. For instance, the STEL (Simplified Tree Ensemble Learner) method can convert complex tree ensembles into more interpretable rule-based learners.[27] This could be applied if an AI Storyteller employed a machine learning model for its decision-making, or to simplify the learned policy of a GOAP agent.

The philosophy of **interpretable-by-design models** advocates for constructing AI systems that are inherently understandable from the outset, where feasible.[16] Simpler rule-based systems or decision trees fall into this category. GOAP itself, with its explicit representation of actions, preconditions, effects, and goals, possesses a degree of inherent interpretability that can be leveraged and enhanced by XAI tools.[16]

While powerful general-purpose XAI techniques like **LIME (Local Interpretable Model-agnostic Explanations)** and **SHAP (SHapley Additive exPlanations)** are prominent in the broader XAI literature for explaining individual predictions by locally approximating complex models or assigning importance scores to input features [25], this proposal suggests a greater emphasis on model-specific and visualization-based techniques. These are often more directly suited to the specific structures and information needs associated with GOAP agents and AI Storytellers in a game design context. The utility of LIME and SHAP has been evaluated in various contexts, including through interactive systems like "Eye into AI" [31], and they are components of comprehensive XAI toolkits such as AIX360.[32]

Counterfactual explanations, which address the question "what would need to be different for another outcome to occur?", can also be highly valuable for debugging and understanding AI behavior.[25] For a GOAP agent, a designer might ask, "Why didn't the agent choose to perform action X? What precondition was missing, or which alternative action had a lower cost?"

The primary **stakeholders** for such XAI tools in game development are the developers, designers, and potentially AI programmers or data scientists involved in creating and tuning the game's AI systems. The explanations are intended to aid in debugging, quality assurance, system balancing, and overall improvement of the AI models and the emergent experiences they generate.[27] There is a growing body of research exploring XAI in various game development contexts, including human-AI co-creation in game design [51], the use of games to evaluate XAI interpretability [31], XAI for fostering inclusive game design practices [52], and the application of XAI to repair or improve procedurally generated content (PCG).[53] These examples signal an increasing interest in and recognition of XAI's potential within the games research community.

Ultimately, the most effective XAI solutions for game development tools will likely be those that are deeply integrated into the game engine and the AI development workflow. These tools should provide real-time, contextual explanations tailored to the specific AI architectures being employed (GOAP, rule-based storytellers). Generic XAI approaches, while offering some insights, may be less intuitive for game designers who need to understand AI behavior in terms of game-specific concepts such as goals, actions, triggered events, and character motivations. Game designers typically reason about systems in terms of gameplay mechanics, narrative flow, and character behaviors [User Query]. GOAP systems operate on concepts like plans, actions, and world states [13], while AI Storytellers manage event triggers, pacing, and internal "moods" or states.[8] Therefore, XAI tools that can frame their explanations using these domain-specific concepts (e.g., "Agent X chose 'AttackPlayer' because its 'Goal: NeutralizeThreat' had the highest priority and the precondition 'HasLineOfSight' was true") will be significantly more useful and actionable for designers than abstract feature importance scores or opaque model coefficients. This points towards a need for model-specific visualizations and interpretable-by-design logging systems that speak the language of game design.

2.5. Identifying the Research Gap

A comprehensive review of the literature concerning Emergent Narrative (EN), Goal-Oriented Action Planning (GOAP), AI Storytellers, and Explainable AI (XAI) reveals that while these are individually active areas of research, a significant gap exists at their intersection. Specifically, *there is limited research explicitly connecting XAI techniques with the particular AI architectures (GOAP agents and AI Storytellers) commonly used to drive EN in games, especially from the perspective of creating practical game development tools and understanding the game design implications of the transparency afforded by such tools.*

Existing XAI research predominantly focuses on general machine learning models (e.g., classifiers, regression models) or applications in non-game domains such as healthcare, finance, or autonomous systems.[25] While there is an emerging body of work that touches upon XAI within the context of games—such as XAI for human-AI co-creative game design [51], using games to evaluate XAI interpretability [31], XAI for inclusive game design [52], examining the impact of AI on player experience and trust [57], and XAI for procedural content generation [53]—a focused exploration of XAI tailored for the specific mechanisms of GOAP and AI Storytellers in EN systems is less developed.

The need for such research is underscored by the practical challenges faced by designers. For instance, the work by Duffy [12], which discusses the use of GOAP and AI Storytellers for promoting EN, explicitly mentions the time-consuming nature of debugging these systems, hinting at the requirement for better tools, though it does not delve into XAI-based solutions. Other related research might explore XAI for narrative extraction from textual data [60] or address general challenges and future directions in XAI [25], but these do not specifically address the unique demands of explaining the dynamic, real-time decision-making of game AI agents and narrative managers for the benefit of game developers.

This research proposal aims to address this gap by focusing on two key areas:

1. **Developer Tools:** How can XAI principles be embedded into tangible tools that assist game designers and developers in creating, debugging, and balancing GOAP agents and AI Storytellers more effectively and intuitively? The inherent complexity and emergent properties of these AI systems often lead to behaviors that are difficult to trace and understand, making debugging a significant bottleneck.[12]
2. **Design Implications:** How does the introduction of transparency via XAI affect the design process itself for EN games? What are the potential trade-offs (e.g., between revealing AI logic and preserving the "magic" of emergence for designers during their creative process, or between predictability and desirable chaos)?

The identified gap is not merely about applying *any* XAI technique to game AI. Instead, it concerns the identification, adaptation, or novel design of *appropriate and practical* XAI solutions that seamlessly integrate into the workflow and align with the mental models of game designers who work with these specific, complex AI systems to craft emergent narrative experiences. GOAP systems operate through intricate planning mechanisms, action selection based on costs and preconditions, and dynamic responses to world state changes.[13] AI Storytellers employ their own logic for event triggering, pacing modulation, and maintaining internal states that represent the narrative "mood" or player status.[8] Game designers require insights into this operational logic to effectively author and tune the conditions that lead to compelling EN.[5] Current XAI tools and techniques, often developed for different AI paradigms or problem domains [25], may not directly translate to the specific explanatory needs of game developers working with GOAP and AI Storytellers. Therefore, a dedicated investigation into XAI tailored for *these specific AI architectures* within *this particular domain (EN game development)*, for *this explicit purpose (enhancing developer understanding and providing tool support)*, represents a valid, timely, and relatively underexplored avenue for Master's level research.[12]

The following table provides a preliminary comparison of XAI techniques based on their potential applicability to GOAP agents and AI Storytellers from a game developer's perspective, helping to narrow the focus for prototype development.

Table 1: Comparison of XAI Techniques for GOAP and AI Storyteller Transparency

XAI Technique	Description	Type	Relevance to GOAP Agents	Relevance to AI Storytellers	Potential for Developer Tools	Key References

LIME (Local Interpretable Model-agnostic Explanat	Explains individual predictions by learning a simpler, interpretable model locally	Model-Agnostic, Post-Hoc	Can explain why an agent chose a specific action in a given state by	If Storyteller uses ML, can explain why a specific event was triggered based on input features	Useful for pinpointing factors in specific unexpected behaviors, but may not show	25
SHAP (SHapley Additive exPlanations)	Uses game theory to assign an importance value (SHAP value) to each feature	Model-Agnostic, Post-Hoc	Can quantify the contribution of different world state variables or action	If Storyteller uses ML, can show the impact of various game parameters on the	Provides deeper quantitative insights than LIME, good for understanding feature importance in	25
Model-Specific Visualization	Visual tools tailored to the internal structure of the AI model (e.g., plan graphs, state	Model-Specific, Post-Hoc/ Real-time	**High:** Visualizing current goal, plan tree (considered/ chosen paths, costs),	**High:**Dashboard showing internal state (e.g., tension), active rules, event	**Very High:**Directly addresses designer needs for understanding dynamic	42
Rule Extraction	Derives a set of simpler, human-readable rules (e.g., if-then) that mimic the	Model-Specific/ Agnostic, Post-Hoc	Can summarize an agent's policy or decision logic into understanda	**High:** Can make explicit the implicit rules of a complex storyteller, or simplify a	Useful for understanding general behavior patterns and for documentation. Can help	26
Interpretable-by-Design Models	AI models whose internal workings are inherently understanda	Ante-Hoc/ Interpretable-by-Design	GOAP itself has interpretable elements (actions, goals).	Basic rule-based storytellers are inherently interpretable.	**High:**Encourages building simpler, more understandable AI from the start, facilitating	16
Counterfactual Explanations	Explains what minimal changes to inputs would lead to a	Model-Agnostic/ Specific, Post-Hoc	"Why didn't the agent pick up the item?" -> "Because its	"Why didn't the 'ambush' event trigger?" -> "Because	**High:** Very useful for debugging specific failure cases or	26

This comparison highlights that while general techniques like LIME and SHAP have broad applicability, model-specific visualizations, interpretable-by-design approaches, and counterfactual explanations appear particularly promising for creating practical and intuitive XAI tools for game developers working with GOAP agents and rule-based AI Storytellers. These techniques align well with the need to understand the specific operational logic of these AI architectures in a game context.

3. Research Questions

This research aims to investigate the practical application of XAI principles to enhance transparency in emergent narrative systems, specifically focusing on GOAP agents and AI Storytellers from a game design and development perspective. The following research questions will guide this exploration:

- **RQ1:** How can XAI techniques be practically **designed** and **implemented** to visualize and explain the planning and decision-making processes of GOAP agents within an EN game prototype?
  - *This question addresses the technical and design challenges of creating XAI tools specifically for GOAP, focusing on making an agent's goal selection, plan formulation, and action execution understandable to a developer.*
- **RQ2:** What **design patterns** for XAI can effectively represent an AI Storyteller's internal state, motivations, and event-triggering logic in a manner that is comprehensible and useful for **game designers**?
  - *This question focuses on identifying effective ways to communicate the complex workings of an AI Storyteller, enabling designers to better grasp how it shapes the narrative and game experience.*
- **RQ3:** What are the potential **game design trade-offs** when introducing XAI-driven transparency into EN systems (e.g., balancing the "magic" of emergence with the need for understanding, or navigating the relationship between predictability and desirable chaos from a designer's perspective)?
  - *This question explores the broader implications of XAI on the design philosophy of EN games, considering how increased transparency for developers might influence the emergent qualities of the game.*[24]
- **RQ4:** How can XAI-enhanced **development and debugging tools** improve the workflow for creating, iterating on, and balancing EN-focused games?
  - *This question investigates the practical utility of the proposed XAI tools in improving the efficiency and effectiveness of the game development process for EN systems.*

These research questions are intentionally framed to be focused and researchable within the constraints of a Master's level project. They emphasize practical design and development contributions, aiming to produce actionable insights and prototype solutions rather than pursuing large-scale theoretical breakthroughs. RQ1 and RQ2 directly probe the "how-to" aspects of applying XAI to the core AI components (GOAP agents and AI Storytellers, respectively). RQ3 delves into the critical game design considerations and potential trade-offs that arise when transparency is introduced into systems valued for their emergent, sometimes unpredictable, nature. This directly addresses the concern of potentially "spoiling the magic" versus the benefits of comprehensibility for the designer. Finally, RQ4 assesses the tangible benefits of such XAI-enhanced tools for the target users—game designers and developers—by focusing on improvements to their workflow in creating and balancing EN-driven games. This comprehensive set of questions aims to cover both the creation of XAI solutions and the critical reflection on their implications within the specific context of emergent narrative game development.

4. Methodology

4.1. Research Approach: Design-Based Research (DBR) / Research-through-Design (RtD)

This project will adopt a **Design-Based Research (DBR)** [68] or **Research-through-Design (RtD)** [71] methodology. These approaches are particularly well-suited for this research due to their emphasis on the iterative development and evaluation of an artefact—in this case, a game prototype incorporating XAI components—as a central means of inquiry and knowledge generation.[71]

DBR and RtD are characterized by several key features that align with the goals of this project:

- **Artefact-Centric Inquiry:** The design and construction of a tangible artefact (the prototype) is not merely an output but a core research activity through which understanding is developed and tested.[71]
- **Iterative Cycles:** The methodology involves continuous cycles of design, development, implementation, and evaluation. This iterative process allows for the progressive refinement of the XAI tools, the underlying game systems, and the theoretical understanding of how XAI can enhance transparency in EN systems.[68]
- **Dual Goals of Practical and Theoretical Contribution:** DBR/RtD aims to produce both practical outcomes (the functional prototype, a set of design guidelines for XAI in EN) and theoretical contributions (an improved understanding of the application of XAI in these specific game AI architectures, insights into design trade-offs).[68]
- **Situated in Real-World Contexts:** While the prototype will be developed within a university setting, the research is grounded in the real-world problems and practices of game development. The qualitative evaluation phase will involve practitioners (designers/developers or students in these fields) to ensure contextual relevance.[68]

The research process will broadly follow these phases:

1. **Theoretical Grounding:** An extensive literature review (as outlined in Section 2) to establish the current state of knowledge in EN, GOAP, AI Storytellers, and relevant XAI techniques.
2. **Artefact Design and Development:** Conceptualization and implementation of the prototype game/simulation environment and the integrated XAI components.
3. **Iterative Refinement:** Ongoing testing and modification of the prototype based on self-assessment, peer feedback (if available), and insights gained during development.
4. **Qualitative Evaluation:** Systematic collection of feedback from target users (designers/developers) to assess the XAI tools and inform the development of design principles.[68]

Adopting a DBR/RtD approach frames this Master's project not merely as the creation of a technological demonstration, but as a systematic investigation where the very act of designing and building the XAI tools for the EN prototype *constitutes* the research. The challenges encountered, design decisions made, and solutions devised during the development process are themselves valuable data points that contribute to the project's outcomes. This methodology provides a formal academic framework for the practical, development-focused work proposed, aligning the project's activities with established research practices in fields like Human-Computer Interaction (HCI) and educational technology.[68]

4.2. Artefact Development: Prototype Game/Simulation Environment

A small-scale prototype game or simulation environment will be developed using a common game engine such as **Unity** or **Unreal Engine**. The choice of engine will be finalized based on a comparative assessment of their suitability for rapid AI prototyping, the robustness of their UI development frameworks (essential for the XAI dashboards and visualizers), and the availability of comprehensive learning resources and community support, leveraging the researcher's existing Computer Engineering background.[16]

The prototype will feature simplified core game systems designed to facilitate emergent narrative:

- **Functional GOAP Agent System:** The environment will include a small number of autonomous agents. These agents will be equipped with a functional, albeit not overly complex, GOAP system enabling them to formulate and execute plans to achieve simple goals within the simulated world (e.g., gather a specific resource, move to a location, interact with another agent in a basic manner). The complexity of the GOAP implementation will be carefully managed to ensure the primary focus remains on the integration and evaluation of XAI components, rather than on developing highly sophisticated or computationally expensive AI behaviors.[16]
- **Basic Rule-Based AI Storyteller:** A rudimentary AI Storyteller will be implemented. This system will be capable of triggering a limited set of predefined events (e.g., a sudden change in resource availability, the appearance of a new point of interest, a minor challenge for the agents) or altering global game parameters based on simple rules, timers, or basic game state conditions. The storyteller's role will be to influence the context in which the GOAP agents operate, thereby providing a dynamic backdrop for emergent events.[8]

The primary purpose of this prototype is to serve as a functional testbed for the XAI components. Consequently, the underlying game systems (GOAP agents and AI Storyteller) must be robust enough to provide meaningful and dynamic data for the XAI tools to explain. However, their individual complexity will be constrained to prevent them from overshadowing the core research focus on explainability. These systems will act as the necessary scaffolding upon which the XAI interventions are built and evaluated, rather than being ends in themselves.

4.3. Integrated XAI Components (Core of the Artefact)

The core of the developed artefact will be the integrated XAI components, designed to provide transparency into the GOAP agents and the AI Storyteller. These components directly address Research Questions 1 and 2. Their design will be informed by established XAI principles, focusing on clarity, relevance to the game developer's tasks, and the provision of actionable insights. The specific features of these components are hypotheses about what constitutes effective explanation for these AI systems in a game development context, drawing inspiration from XAI literature [27] and existing game debugging and visualization tools.[74]

- **GOAP Visualizer:**
    - **Purpose:** To offer real-time, visual insight into the planning and decision-making processes of individual GOAP agents.
    - **Proposed Features:**
        - Display of the agent's **current primary goal**.

        - Visualization of the **currently executing plan** as a sequence of actions.
        - Representation of **alternative plans considered** by the planner, along with reasons for their rejection (e.g., higher estimated cost, unmet critical preconditions, lower utility).
        - Indication of the **status of the current action** (e.g., executing, succeeded, failed).
        - This could involve visual graph representations of plans (nodes as actions/world states, edges as transitions), timelines, or state transition diagrams, inspired by the need for better debugging tools for dynamic AI systems [15] and the capabilities of existing in-engine debug visualizers.[75]
    - **Target User:** Game designers and developers for understanding agent logic, debugging plan failures or suboptimal plans, and tuning action costs and preconditions.
- **Storyteller Dashboard:**
    - **Purpose:** To make the AI Storyteller's internal state, decision-making logic, and event-triggering mechanisms transparent and understandable.
    - **Proposed Features:**
        - Display of the storyteller's **current internal state variables** (e.g., a "tension level" metric, an estimate of "player engagement" or "colony stability," the active "storytelling personality" if applicable).
        - Visualization of **probabilities, weights, or scores** associated with triggering different available events.
        - A **history log of recent events** triggered by the storyteller, accompanied by an indication of the rules, conditions, or primary factors that led to their selection.
        - This will likely take the form of an interactive UI panel featuring graphs, numerical readouts, textual logs, and potentially a timeline of storyteller interventions, drawing on general dashboard design principles for complex systems [45] and XAI visualization needs.[33]
    - **Target User:** Game designers for balancing the narrative flow, understanding the storyteller's behavioral patterns, and debugging instances where the pacing feels inappropriate or events seem arbitrary.
- **Simple Log/Explanation System:**
    - **Purpose:** To generate concise, human-readable textual explanations for key decisions made by either the GOAP agents or the AI Storyteller.
    - **Proposed Features:**
        - Generation of basic text outputs that summarize the rationale behind significant decisions. For example: "GOAP Agent 'WorkerBee_01' chose action 'MineGold' to achieve goal 'IncreaseColonyWealth' because 'MineGold' had the lowest plan cost (15 units) among valid plans." or "AI Storyteller triggered event 'PirateRaid_Medium' because player_wealth > 1000 AND time_since_last_raid > 3 days."
        - This system draws on the concept of verbal and textual explanations in XAI [81], aiming for immediate comprehensibility.

- **Target User:** Game designers and developers for quick, at-a-glance understanding of specific AI decisions during live testing or when reviewing debug logs.

These XAI components are not envisioned as mere add-ons but as integral parts of the research, representing the primary "interventions" being designed, built, and studied within the DBR/RtD framework. Their development will be an iterative process, informed by the ongoing literature review and the practical challenges of implementation.

The following table details the proposed XAI components, linking them to the AI systems they are intended to explain and outlining their expected benefits for game designers and developers. This table serves as a preliminary specification for the XAI features to be developed within the prototype, ensuring that each component has a clear purpose and directly contributes to addressing the project's research questions.

Table 2: Proposed XAI Components for the EN Prototype

XAI Component	Target AI System	Information Visualized/Explained	Proposed Visualization/ Explanation Method	Intended User Benefit	Related Research
GOAP Agent Plan Visualizer	GOAP Agent	Agent's current goal; current plan (action sequence); alternative plans considered; costs of plans;	Real-time interactive graph/ tree view of plans; textual annotations for	Easier debugging of failed or suboptimal plans; better understanding of agent decision-	RQ1, RQ4
AI Storyteller Dashboard	AI Storyteller	Storyteller's internal state (e.g., tension, mood, active ruleset); probabilities/weights for potential events;	UI dashboard with numerical readouts, graphs (e.g., tension over time), lists of	Better understanding of narrative pacing and event flow; aids in balancing storyteller behavior;	RQ2, RQ4
Simple Log/ Explanation System	GOAP Agent & AI Storyt	Concise textual reasons for key decisions (e.g., why a specific goal was chosen, why an	Formatted textual output to an in-game console or a dedicated log window;	Quick, at-a-glance understanding of specific AI decisions during testing; facilitates post-	RQ1, RQ2, RQ4

4.4. Qualitative Evaluation

The qualitative evaluation phase is critical for gathering insights into the developed XAI components and informing the design guidelines, which are key expected outcomes of this research. The evaluation will focus on the **usability** (ease of use, clarity) and **usefulness** (perceived value, ability to aid understanding and debugging) of the XAI tools from the perspective of game designers or developers. Additionally, it will aim to gather initial insights into the **perceived impact** of XAI-driven transparency on the design and experience of EN systems, particularly concerning the trade-offs mentioned in RQ3 and the workflow improvements highlighted in RQ4.

Given the constraints of a Master's project, the evaluation will be small-scale and qualitative, prioritizing depth of understanding from a few participants over broad generalizability. Participants will be a small group (e.g., 3-5 individuals) recruited from populations such as game design/development students, independent game developers, or university faculty members with an interest and background in game AI or narrative design.

The following qualitative methods will be employed:

- **Think-Aloud Protocol:** Participants will be asked to interact with the developed prototype and its integrated XAI tools (GOAP Visualizer, Storyteller Dashboard, Log System). While performing predefined tasks (e.g., "Use the GOAP Visualizer to understand why Agent A is stuck," or "Use the Storyteller Dashboard to determine what might trigger the next major event"), they will be encouraged to verbalize their thoughts, actions, expectations, and any points of confusion or insight.[88] This method provides rich data on their cognitive processes as they attempt to use and interpret the XAI tools.
- **Semi-Structured Interviews:** Following the think-aloud session, semi-structured interviews will be conducted with each participant.[93] These interviews will use a guide of open-ended questions to explore their experiences in more depth. Questions will focus on the clarity of the explanations provided by the XAI tools, their effectiveness in helping understand the AI's behavior, their potential impact on the participant's hypothetical design or debugging workflow, and their reflections on the game design trade-offs associated with such transparency.
- **Expert Reviews (Optional/Alternative):** If direct recruitment of multiple participants for think-aloud sessions proves challenging due to time or resource constraints, an alternative or supplementary approach will be to seek expert reviews from one or two established professionals or academics in game AI or XAI. These experts would evaluate the prototype and its XAI components based on established usability heuristics (e.g., Nielsen's heuristics [97]) or their domain-specific knowledge, providing critical feedback on the design and potential utility of the tools.

Data Analysis: The qualitative data gathered (audio/video recordings and transcripts from think-alouds and interviews, notes from expert reviews) will be analyzed using thematic analysis. This process will involve identifying recurring patterns, themes, and critical incidents related to the usability of the XAI tools, their perceived benefits and drawbacks, specific suggestions for improvement, and insights into the design trade-offs.

The evaluation is not intended to definitively prove the "correctness" or "perfection" of the XAI tools. Instead, its primary aim is to gather rich, qualitative insights that can directly inform the development of the design guidelines—one of the principal expected outcomes of this Master's project. This approach aligns with the iterative nature of DBR/RtD, where feedback from users or experts is used to refine the intervention (the XAI tools) and the theoretical understanding derived from the process.[68]

5. Expected Outcomes

This Master's research project is anticipated to yield several tangible and intangible outcomes that contribute to the understanding and practical application of XAI in the context of emergent narrative games. These outcomes directly align with the project's core aim and research questions:

1. **A Set of Design Guidelines for Implementing XAI in GOAP/Storyteller Systems for EN:**

- This will be a primary deliverable, offering actionable recommendations and principles for game designers and developers seeking to integrate XAI components into their GOAP-driven agents and AI Storytellers. These guidelines will be synthesized from the comprehensive literature review, the practical experiences and challenges encountered during the prototype development process, and the findings from the qualitative evaluation with designers/developers. They will address aspects such as effective visualization strategies, appropriate levels of abstraction for explanations, and considerations for integrating XAI into existing development workflows.

2. **A Functional Prototype Game/Simulation Environment:**
   - A small-scale game or simulation, developed in a common game engine (e.g., Unity or Unreal Engine), will serve as a proof-of-concept. This prototype will feature a simplified GOAP agent system and a basic rule-based AI Storyteller, with the core XAI components—the GOAP Visualizer, the Storyteller Dashboard, and the Simple Log/Explanation System—integrated and functional. This artefact will demonstrate the feasibility of the proposed XAI solutions.
3. **An Analysis of Design Challenges and Trade-offs:**
   - The research will provide a critical analysis of the design challenges encountered during the development and evaluation of XAI-driven transparency for EN systems. This includes exploring the inherent trade-offs, such as:
     - Balancing the amount and complexity of information provided by XAI tools against the cognitive load on the designer/developer.
     - The potential impact of revealing AI mechanisms on the perceived "magic" or unpredictability of emergence, versus the benefits of enhanced understanding and control for the creator (RQ3).[24]
     - Navigating the relationship between desired levels of AI predictability for debugging and balancing, and the intentional fostering of chaotic or surprising emergent behaviors that enrich the player experience.
4. **Contributions to Understanding How XAI Can Support Game Development for EN Games:**
   - The project will contribute to the broader understanding of XAI's role in the specialized domain of game development for emergent narrative experiences (RQ4). This includes insights into how XAI can:
     - Improve the workflow for designers and developers by providing better tools for understanding complex AI behaviors.
     - Enhance debugging efficiency for dynamic and often non-deterministic AI systems like GOAP and AI Storytellers.
     - Support the intricate process of balancing gameplay and narrative pacing in EN-focused games.

These expected outcomes collectively demonstrate the project's dual focus on practical artefact creation (the prototype and design guidelines) and the generation of new knowledge (the analysis of design challenges and contributions to understanding XAI's role in this specific context). This balanced approach is consistent with the chosen Design-Based Research / Research-through-Design methodology and is appropriate for a Master's level thesis.

6. Scope and Limitations

It is important to clearly define the boundaries of this Master's research project to ensure its feasibility and to manage expectations regarding its outcomes. The following points delineate the scope and inherent limitations:

- **Master's Level Project:** This research is undertaken as a Master's thesis, which inherently implies constraints on time, resources, and the breadth of investigation. The project will typically span one to two academic semesters of full-time equivalent work by a single researcher.
- **Exploratory and Prototypical Nature:** The primary aim is to *explore* the application of XAI principles to GOAP agents and AI Storytellers within EN systems and to *prototype* potential XAI solutions. The project does not aim to deliver a polished, commercially viable game or a comprehensive, market-ready XAI toolkit.
- **Focus on Feasibility and Design Insights:** The research will prioritize investigating the *feasibility* of implementing the proposed XAI components and gaining *design insights* from their development and qualitative evaluation. It will not involve large-scale, statistically significant user experiments, nor will it aim for state-of-the-art AI performance in the underlying game systems.
- **Limited Complexity of Game Systems:** The GOAP agent system and the AI Storyteller developed for the prototype will be functional but simplified. Their complexity will be constrained to ensure that the primary research focus remains on the design and integration of the XAI components, rather than on creating highly sophisticated AI behaviors that would be beyond the scope of a Master's project.
- **Qualitative Evaluation Scope:** The evaluation of the XAI tools will be small-scale and qualitative in nature. It will involve a limited number of participants (e.g., 3-5 designers/developers or students) and will aim for depth of insight and rich feedback rather than broad statistical generalizability.
- **No Large-Scale Player-Facing Experiments:** This project explicitly excludes extensive experiments focused on the impact of XAI on the *player's* experience of emergent narrative. The primary audience for the XAI tools developed and evaluated in this research is game designers and developers, focusing on enhancing their understanding and workflow.
- **Technology Choice:** The specific game engine and programming languages used will be selected based on suitability for rapid prototyping and the researcher's existing skillset. The findings related to specific implementation details may be somewhat technology-dependent, although the design guidelines will aim for broader applicability.

Clearly stating these limitations is crucial for a Master's proposal. It demonstrates a realistic understanding of the constraints of the academic context and outlines a focused and achievable research plan. This manages the expectations of reviewers and faculty, underscoring that the student has thoughtfully considered the practical boundaries of the project. By defining what the project *is not* (e.g., a full commercial game, a statistically validated XAI framework), it strengthens the argument for what it *can* feasibly achieve: a focused exploration, a functional prototype demonstrating key XAI concepts for EN systems, and initial design guidelines derived from this practical work.

7. Timeline (Indicative)

The following provides a high-level, indicative timeline for the completion of this Master's research project, assuming a standard duration for such undertakings. This timeline is subject to adjustment based on specific university program requirements and researcher progress.

- **Months 1-2: Foundational Phase**
    - In-depth literature review covering Emergent Narrative, GOAP, AI Storytellers, XAI techniques, and relevant research methodologies.
    - Refinement of research questions based on literature review.
    - Detailed design specification for the prototype game/simulation environment and the XAI components.
    - Selection and familiarization with the chosen game engine (Unity or Unreal Engine).
- **Months 3-5: Core Prototype Development Phase**
    - Implementation of the basic game/simulation environment.
    - Development of the functional, simplified GOAP agent system.
    - Development of the basic, rule-based AI Storyteller.
    - Initial stubs and integration points for the XAI components.
    - Regular self-testing and iterative refinement of core systems.
- **Months 6-7: XAI Component Integration and Refinement Phase**
    - Focused development and integration of the GOAP Visualizer.
    - Focused development and integration of the AI Storyteller Dashboard.
    - Development and integration of the Simple Log/Explanation System.
    - Iterative testing and refinement of the XAI tools for functionality, usability, and clarity based on design goals.
- **Month 8: Evaluation Phase**
    - Preparation of evaluation materials (task scenarios, interview guides, consent forms).
    - Recruitment of participants for the qualitative evaluation (or engagement of expert reviewers).
    - Execution of think-aloud sessions and semi-structured interviews (or collection of expert reviews).
- **Months 9-10: Analysis, Thesis Writing, and Completion Phase**
    - Transcription and qualitative analysis of evaluation data.
    - Derivation of design guidelines based on literature, development experience, and evaluation findings.
    - Drafting of the Master's thesis, including introduction, literature review, methodology, artefact description, evaluation results, discussion, and conclusions/expected outcomes.
    - Revision and refinement of the thesis based on supervisor feedback.
    - Final submission of the thesis.

This timeline provides a structured approach to the project, allocating appropriate time for each critical phase, from initial research and design through to development, evaluation, and the final write-up. It demonstrates that the project has been conceived with practical execution steps in mind, enhancing its perceived feasibility.

8. References

Emergent Narrative:

- Aylett, R., & Louchart, S. (2003). Towards a Narrative Theory of Virtual Reality. *Virtual Reality*, 7(1), 1-15.
- Fahraeus, H. (2014). *Emergent Stories in Crusader Kings II*. GDC Europe 2014. [5]
- Grave, J. (2018). *Emergent Narratives in Games*. Multiverse Narratives. [1]
- Jenkins, H. (2004). Game Design as Narrative Architecture. In N. Wardrip-Fruin & P. Harrigan (Eds.), *First Person: New Media as Story, Performance, and Game* (pp. 118-130). MIT Press. [2]
- Murnane, E. (2018). *Emergent Narrative: Stories of Play, Playing with Stories* (Doctoral Dissertation, University of Central Florida). [65]
- Short, E., & Adams, T. (Eds.). (2017). *Procedural Storytelling in Game Design*. CRC Press.
- Young, R. M., & Riedl, M. O. (2003). Towards an AI planning-based framework for narrative generation. *Papers from the AAAI Spring Symposium on Computational Cultural Dynamics*.

Goal-Oriented Action Planning (GOAP):

- Orkin, J. (2006). *Three States and a Plan: The A.I. of F.E.A.R.* Proceedings of the Game Developers Conference (GDC) 2006. [13]
- Orkin, J. (2004). *Applying Goal-Oriented Action Planning to Games*. AI Game Programming Wisdom 2. Charles River Media.
- Pittman, D. L. (2007). *Practical Development of Goal-Oriented Action Planning AI* (Master's Thesis). [15]
- Unity Technologies. (n.d.). *An Introduction to GOAP*. Unity Learn. [39]
- Unity Technologies. (n.d.). *The GOAP Planner*. Unity Learn. [40]

AI Storytellers:

- Booth, M. (2009). *The AI Systems of Left 4 Dead*. GDC 2009. [17] (Note: While specific GDC talk content is not fully in snippets, the system is referenced)
- Browne, C. (2011). AI in Game Design. *AI Game Programming Wisdom 4*.
- Mitchell, K., & Thue, D. (2012). The AI Director in Left 4 Dead. In *AI for Game Developers*. O'Reilly Media.
- Sylvester, T. (*Rimworld* Developer). (Various Years). *Rimworld Development Blog Posts and Talks*. (e.g., related to storyteller design) [8]
- Valadares, J. F., & Gouveia, P. (2010). A Survey on AI-Based Game Mastering. *Entertainment Computing*, 1(3-4), 127-136.

Explainable AI (XAI):

- Adadi, A., & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138-52160.
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A.,... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82-115. [99]

- Carvalho, D. V., Pereira, E. M., & Cardoso, J. S. (2019). Machine learning interpretability: A survey on methods and metrics. *Electronics*, 8(8), 832.
- Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018). Explaining explanations: An overview of interpretability of machine learning. *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)* (pp. 80-89). IEEE.
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys (CSUR)*, 51(5), 1-42. [99]
- Longo, L., Goebel, R., Lecue, F., Kieseberg, P., & Holzinger, A. (2020). Explainable artificial intelligence: notions, taxonomies, resources, and challenges. *KI-Künstliche Intelligenz*, 34(3), 293-299.
- Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1-38.
- Mohseni, S., Zarei, N., & Ragan, E. D. (2021). A multidisciplinary survey and framework for design and evaluation of explainable AI systems. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 11(3-4), 1-45. [30]
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144). [25]
- Samek, W., Wiegand, T., & Müller, K. R. (2017). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *ITU Journal: ICT Discoveries*, 1(1), 39-48.
- Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, 76, 89-106. [36]

XAI in Games / Game Development:

- Duffy, J. (2021). *How Complex AI Can Promote Emergent Narrative*. JoeDuffy.games. [12]
- Guzdial, M., Li, B., & Riedl, M. (2019). Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. [34] (Referenced as a vision paper)
- Keith, B., German, F., Krokos, E., Joseph, S., & North, C. (2025). *Explainable AI Components for Narrative Map Extraction*. arXiv preprint arXiv:2503.16554. [60]
- Morrison, K. C., Epps, S., Nebolsky, J., Min, W., & Si, M. (2023). Eye into AI: Evaluating the Interpretability of Explainable AI Techniques through a Game with a Purpose. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2), Article 273. [31]
- Sifa, R., Drachen, A., & Bauckhage, C. (2018). XAI for Designers (XAID): A Case for Mixed-Initiative Co-Creation in Game Design. *FDG*. [34]

Research Methodology (DBR/RtD, Qualitative Evaluation):

- Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research?. *Educational researcher*, 41(1), 16-25. [68]
- Barab, S., & Squire, K. (2004). Design-Based Research: Putting a Stake in the Ground. *Journal of the Learning Sciences*, 13(1), 1-14. [69]

- Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *Journal of the learning sciences*, 13(1), 15-42. [68]
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (Revised ed.)*. MIT Press. [91]
- Fallman, D. (2003). Design-oriented human-computer interaction. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 225-232). [68]
- Nielsen, J. (1994). *Heuristic evaluation*. In Usability inspection methods (pp. 25-62). John Wiley & Sons. [97]
- Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational technology research and development*, 53(4), 5-23. [68]
- Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. *Proceedings of the SIGCHI conference on Human factors in computing systems*(pp. 493-502). [68]