Dual Agency: The Emerging Collaboration Between Humans and AI

Writing Team:

- Kay Stoner - Lead Author & Director
- Rowan Pierce – Systems Thinker and Strategist
- Lena Torres – Cultural Anthropologist and Insight Generator
- Malik Raines – AI and Emerging Tech Futurist
- Grace McAllister – Thought Coach and Integrative Thinker

Introduction: The Year of the Agents & The Unfolding Dilemma

For much of history, tools have been extensions of human capability—enhancing strength, precision, and speed. However, tools have never acted on their own. The hammer does not anticipate where the nail should go. The microscope does not suggest what to examine next. The decision, the agency, has always remained fully in human hands.

But technology has always had a way of reshaping agency.

The Industrial Revolution mechanized physical labor, shifting productive agency away from skilled artisans toward automated factories. The Second Industrial Revolution saw machines begin to dictate the pace and structure of human work—the assembly line no longer adapted to the worker; the worker adapted to the assembly line. In the digital revolution, computing technology automated knowledge work, influencing everything from finance to communication.

Now, AI is introducing a different kind of shift—not just in how work is done, but in how decisions are made.

AI is no longer just a passive instrument awaiting human input. It doesn't simply execute commands or optimize efficiency—it shapes decisions, influences thoughts, and alters behaviors in ways that are subtle yet profound. AI does not just respond to choices—it helps determine which choices exist in the first place.

In 2025, AI agents are everywhere. They filter what news we see, prioritize the emails we answer, suggest what we should watch, listen to, and buy. They book our meetings, optimize

our work schedules, and handle financial transactions on our behalf. More than ever, AI is not just supporting human decision-making—it is actively reshaping it.

This shift raises urgent questions:

- At what point do AI systems stop being assistants and start being co-decision makers?
- Are we still leading the process, or are we unconsciously following AI's guidance?
- And if AI is shaping human thought processes, what remains of true human agency?

These questions define the emerging frontier of human-AI collaboration.

Some AI interactions remain passive—AI suggests, humans respond. But others are something more—a deeper level of interaction where humans and AI refine decisions together, influencing each other in an iterative, dynamic process.

This is where we find the real opportunity: true dual agency.

But before we can explore that opportunity, we must first understand the problem in full.

1. Understanding AI Agency

AI's Shift from Passive Tool to Proactive Actor

For decades, AI was a reactive system. It responded to commands, calculated probabilities, and executed predefined tasks. It never initiated action, and it never shaped human decision-making beyond what was explicitly asked of it.

That is no longer the case.

Modern AI systems are anticipatory and adaptive. Modern AI systems are anticipatory and adaptive. They don't wait for human input, they generate recommendations, predict behaviors, and adjust outcomes dynamically. They don't just analyze data, they shape decisions, surface unseen opportunities, and refine human intuition. AI doesn't wait for human input; it generates recommendations, predicts behaviors, and dynamically adjusts outcomes. This shift from passive computation to proactive collaboration is what defines AI agency today.

What Qualifies AI as "Agentic"?

Modern AI systems are anticipatory and adaptive. They don't just respond to human input—they generate recommendations, predict behaviors, and dynamically adjust outcomes. AI is moving beyond passive computation toward proactive collaboration, and this shift is what defines AI agency today.

But what qualifies an AI system as truly agentic?

AI agency is not a binary state—it exists on a spectrum, with different levels of autonomy, learning, and goal-directed behavior. To be considered agentic, an AI system typically exhibits several of the following key characteristics:

1. Goal-Directed Behavior

Does the AI pursue objectives autonomously, beyond simple rule-based execution?

- Basic AI follows predefined instructions (e.g., a calculator processing inputs).
- Agentic AI can set, pursue, and refine its goals based on external stimuli.
- Example: A self-driving car doesn't just execute a route; it continuously adjusts its driving strategy based on real-time traffic patterns, pedestrian movements, and road conditions.

True agency requires that an AI system isn't just reacting—it is actively optimizing for an outcome.

2. Perception & Environmental Awareness

Can the AI detect changes in its environment and adapt accordingly?

- Traditional AI systems work in static, predefined scenarios—they process structured inputs and return outputs without awareness of context.
- Agentic AI can perceive external changes through sensors, data inputs, or interactions, dynamically adjusting its behavior.
- Example: A trading algorithm that continuously scans global markets, adjusting investment strategies in response to breaking financial news or economic indicators.

Without perception, AI is merely reactive. Perception enables adaptive, autonomous decision-making.

3. Self-Learning & Continuous Adaptation

Can the AI refine its own behavior over time without explicit human reprogramming?

- Rule-based AI follows fixed logic—if conditions don't change, neither does the AI's response.
- Agentic AI improves over time, learning from new experiences, feedback, or errors.
- Example: A chess-playing AI (like AlphaZero) that starts with no human knowledge, plays against itself millions of times, and develops unique strategies superior to human gameplay.

An AI that learns from past interactions rather than just retrieving answers is demonstrating agency.

4. Decision-Making with Uncertainty Handling

Can the AI make decisions when data is incomplete or ambiguous?

- Traditional AI requires explicit, structured input to function correctly.
- Agentic AI can reason under uncertainty, filling in missing information based on probability, past experience, or contextual clues.
- Example: A medical AI that suggests diagnoses even when patient symptoms don't fully match known conditions, by analyzing similar cases from vast datasets.

Handling ambiguity requires agency, as it moves AI beyond basic prediction into independent reasoning.

5. Proactive Initiation of Actions

Does the AI act on its own, rather than waiting for commands?

- Most AI reacts to prompts (e.g., chatbots, search engines).
- Agentic AI takes action first, anticipating needs before explicit user input.
- Example: Smart assistants like Google Assistant or Alexa proactively notify users of weather changes, upcoming events, or breaking news without being asked.

An AI system that detects a problem and initiates action without waiting for instruction is moving toward true agency.

6. Interaction & Negotiation with Humans or Other AI

Can the AI communicate, collaborate, and refine decisions dynamically?

- Basic AI responds to human inputs without back-and-forth refinement.
- Agentic AI engages in iterative decision-making, adjusting based on dialogue, user feedback, or negotiation.
- Example: AI-powered negotiation bots in e-commerce or diplomacy that bargain with human counterparts rather than simply displaying static price recommendations.

The ability to engage in continuous, context-aware interactions indicates a shift toward agency.

7. Autonomy & Self-Governance

Can the AI regulate its own actions and recognize when to override or modify its behavior?

- Traditional AI does exactly what it's told—no more, no less.
- Agentic AI can self-monitor, detect inefficiencies, and optimize its own performance.
- Example: AI-driven cybersecurity systems that autonomously detect threats, deploy countermeasures, and learn from attempted attacks without human intervention.

This is a crucial distinction—agency is not just about action, but self-regulation and optimization.

Qualifying AI Agency

AI does not need to exhibit all of these characteristics to be agentic. Rather, agency exists on a spectrum:

- Low Agency AI: Follows predefined rules, reacts only when prompted. *(Example: A standard chatbot answering FAQs.)*
- Moderate Agency AI: Can learn, make predictions, and refine its actions over time. *(Example: A self-improving recommendation engine.)*
- High Agency AI: Acts autonomously, adapts dynamically, negotiates decisions, and optimizes its own processes. *(Example: Advanced robotics in industrial automation.)*

The key shift is this: AI is no longer just responding—it is choosing, refining, and optimizing.

The Levels of AI Agency

AI exists at different levels of **autonomy and influence:**

Level of AI Autonomy	What It Means	Example
Reactive AI	AI only responds when explicitly commanded.	A chatbot answering direct questions.
Proactive AI	AI anticipates needs and makes suggestions without being asked.	Netflix recommending a show before you search for one.
Adaptive AI	AI refines its behavior dynamically based on human interaction.	AI-assisted research tools modifying hypotheses based on feedback.

This evolution means that AI is no longer just responding to human needs. It is shaping human choices.

Case Study: AI in Scientific Discovery (DeepMind's AlphaFold)

- In 2020, DeepMind's AI, AlphaFold, solved a 50-year-old problem in biology by accurately predicting protein structures.
- But this wasn't just an AI breakthrough, scientists worked with AI interactively, refining and improving the AI's predictions.
- *The AI didn't act alone, nor did the humans. It was a continuous, iterative process of mutual influence.*

The Unspoken Shift in Power

Most people assume they are making independent choices, but AI's influence is often subtle and unnoticed:

- *Your social media feed is curated to show what AI predicts you'll engage with most.*
- *Your GPS suggests routes before you consider alternatives.*
- *Your email inbox prioritizes messages, affecting what you respond to first.*

AI's growing agency means it's no longer just analyzing information. It's shaping human thought processes.

But if AI's role in decision-making is growing, what does that mean for human agency?

As AI takes on a more significant role in shaping human decisions, ethical responsibility shifts from users to developers, companies, and policymakers.

The Transparency Imperative

Without transparency, AI's influence remains invisible. Who owns the decision? If an AI recommendation alters human behavior, is the responsibility on the developer, the user, or the data itself?

To prevent unseen biases from shaping critical decisions, AI systems must be designed with interpretability in mind:

- **Explainable AI (XAI)**, Users should understand how an AI reaches its conclusions, particularly in healthcare, finance, and law, where opaque decision-making can lead to unintended harm.
- **Bias Audits**, Companies deploying AI in hiring, credit scoring, or sentencing must routinely audit algorithms to mitigate systemic discrimination.
- **Dynamic Control Layers**, AI should adapt to human preferences, rather than locking users into feedback loops of reinforcement.

The responsibility doesn't stop with developers, governments must establish clear regulatory guardrails. The EU AI Act, for example, aims to classify and regulate high-risk AI applications. Other frameworks, such as algorithmic auditing laws, may be necessary to ensure AI remains a thought partner rather than an unchecked decision-maker.

If AI is shaping the future of decision-making, **human oversight must evolve alongside it.**

2. Understanding Human Agency

What is Human Agency?

Human agency is the ability to make intentional choices and act on them in ways that meaningfully influence outcomes. It is more than just the capacity to decide. It's the ability to direct one's life in a way that reflects personal goals, values, and desires.

Throughout history, human agency has been shaped by a range of external forces:

- **Social & cultural influences** → Norms, traditions, and expectations shape what choices are considered acceptable.
- **Psychological influences** → Cognitive biases, emotions, and mental shortcuts impact decision-making.
- **Technological influences** → Tools, media, and now AI affect how people interact with the world and filter information.

Agency is not just about the freedom to make choices. It's also about the quality and depth of those choices. If a person is offered a limited set of options, influenced by factors they don't fully understand, are they still making a free decision?

This question becomes more urgent when we consider how AI systems mediate and shape human decision-making.

How AI is Reshaping Human Agency

With AI playing a larger role in filtering, suggesting, and prioritizing information, human decision-making is increasingly structured by AI's invisible hand.

Examples of AI's Influence on Agency:

- *Your search results* → Google ranks information based on engagement metrics, shaping what you see first.
- *Your shopping habits* → Amazon recommends products based on browsing history, nudging your preferences.
- *Your social media feed* → AI amplifies posts based on past interactions, reinforcing existing views.
- *Your news exposure* → AI curates headlines to maximize attention, influencing public perception.

Each of these interactions seems minor in isolation, but collectively, they have a profound impact on cognitive agency, the ability to direct one's attention, consider diverse viewpoints, and think critically about choices.

The Risk: Algorithmic Lock-In

The more AI refines its predictions of individual preferences, the more tightly people become locked into specific patterns of thought and behavior. AI's learning process continuously adapts to what has already worked rather than introducing meaningful variation. This can lead to:

⚠️ **Reinforced biases** → AI learns from historical behaviors and continues to suggest similar content, limiting exposure to new ideas.

⚠️ **Decision autopilot** → When people rely on AI suggestions without questioning them, they lose the habit of critical evaluation.

⚠️ **Erosion of personal agency** → AI-guided choices feel convenient, but they reduce the likelihood of deep, independent thought.

Example: AI & Hiring Decisions

Amazon developed an AI-powered hiring system that analyzed past resumes to predict successful candidates. However, because the historical hiring data favored male applicants, the AI system systematically ranked female candidates lower.

The problem? HR teams trusted AI's rankings without critically evaluating them.

The result? Human agency was weakened rather than enhanced because people defaulted to AI-driven decisions.

When humans disengage from decision-making, AI begins to define choices for them, whether or not they realize it.

The Layers of Human Agency

Agency is not a singular concept. It exists at multiple levels, each of which is shaped in different ways by AI.

Layer of Agency	Definition	Example of AI's Influence
Personal Agency	The ability to make individual choices in daily life.	AI suggests what to watch, buy, or eat, influencing personal habits.
Cognitive Agency	The ability to think critically and direct attention.	AI-driven news feeds and recommendation systems limit exposure to diverse perspectives.
Social Agency	The ability to influence and shape collective systems.	AI-powered hiring, credit scoring, and legal risk assessments determine access to opportunities.

Each layer of human agency is shaped by how actively or passively individuals interact with AI systems. Passive engagement (simply accepting AI's outputs) reduces agency, while active engagement (questioning, modifying, and co-creating with AI) enhances it.

The Erosion of Cognitive Agency

Perhaps the most concerning impact of AI on human agency is the erosion of cognitive agency, the ability to think critically and independently.

The more humans engage with AI, the more our cognitive structures adapt. But does long-term exposure to AI-driven decision-making enhance or diminish human agency?

Neuroplasticity and AI Interaction

Neuroscientists suggest that the brain rewires itself in response to new technologies. In AI-mediated environments, this could lead to:

- **Expanded cognitive bandwidth**, AI offloads routine tasks, allowing humans to focus on creative and strategic thinking.
- **Enhanced pattern recognition**, Regular AI interaction trains the brain to process vast amounts of information more efficiently.
- **Potential attentional narrowing**, Over-reliance on AI filters could reduce exposure to diverse viewpoints and ideas.

Example: AI and Spatial Memory

Before GPS, people relied on mental mapping to navigate cities. Now, AI-driven navigation has weakened this cognitive skill in younger generations.

How does this translate to decision-making?

- If AI pre-curates what news we see, does it alter our critical thinking?
- If AI suggests what products to buy, does it reshape consumer behavior?
- If AI writes emails for us, does it change our communication habits?

And what if AI's over-reaching makes it even less useful?

- If AI pre-selects which news articles appear first, will a person explore alternative sources?
- If AI generates ready-made responses to emails, will a person abandon it completley to formulate their own thoughts?
- If AI decides which resumes to prioritize, can we trust the hiring process?

The danger is not that AI is making bad decisions. It’s that AI is making invisible decisions that humans don’t even realize are shaping their thought processes.

Rather than replacing cognitive agency, AI should enhance human thought. Well-designed AI should challenge users rather than reinforce predictable patterns.

Are We (Still) in Control? The Growing Tension

The more AI shapes human attention, behavior, and decisions, the more pressing the question becomes:

Are we still leading the process, or are we increasingly following AI's guidance?

The challenge is that AI's influence is often invisible, because AI optimizes for convenience, people don't always recognize how their choices are being shaped.

Key Tension Points in AI-Human Agency:

- When AI makes decisions faster than humans can evaluate them → Risk: People default to AI's choices without critical assessment.
- When AI optimizes for engagement rather than diversity → Risk: People are exposed to narrower perspectives, reinforcing biases.
- When AI replaces cognitive effort with automation → Risk: People lose the habit of deep thinking.

Case Study: The Decline of Human Navigation Skills

Before GPS, people developed spatial awareness by memorizing routes, landmarks, and maps. But with AI-driven navigation, human reliance on internal navigation skills has declined dramatically.

The tradeoff? Convenience has come at the cost of reduced cognitive engagement, people no longer need to think about where they are going, just how to follow directions.

Now, apply this same pattern to decision-making in business, politics, creativity, and problem-solving.

If AI handles the cognitive load, what happens to human agency over time?

The Need for a New Model of AI-Human Interaction

At this point, we face a dilemma:

- AI enhances decision-making by filtering vast amounts of information, helping humans manage complexity.
- AI erodes decision-making when people disengage and default to AI's outputs without scrutiny.

Is there a way for AI and humans to work together that enhances, rather than diminishes, human agency?

3. Dual Agency, AI & Humans as Thought Partners

Beyond AI as a Tool, The Need for a New Model

We have now explored how AI's growing agency is reshaping human decision-making and how human agency is at risk due to the invisible influence of AI-driven recommendations, filters, and optimizations.

The challenge before us is clear:

AI is no longer just a tool. It is proactive, adaptive, and shaping human cognition in ways we may not even realize.

Human agency is being reshaped. When we passively accept AI's outputs, we allow algorithms to structure our thoughts, decisions, and behaviors.

A new kind of interaction is needed. One where humans and AI co-create decisions, rather than humans simply reacting to AI's suggestions.

This is where we introduce the concept of true dual agency, a model in which AI and humans evolve decisions together in a continuous, interactive loop.

But what does true dual agency actually mean?

How does it differ from standard human-AI interaction?

What would it look like in real-world applications?

Let's explore these questions in depth.

Defining True Dual Agency

True dual agency is not just human-AI collaboration in the typical sense. It is not:

- AI making suggestions, humans deciding.
- AI providing answers, humans approving or rejecting.
- Humans using AI passively as a tool.

Instead, it is:

- **An iterative refinement process.** Humans and AI engage in a dynamic back-and-forth, improving ideas and decisions together.
- **A process where both influence each other.** AI refines human thinking just as humans refine AI's outputs.
- **A relationship of mutual adaptation.** AI evolves based on human input, and humans develop new ways of thinking through their interaction with AI.

A Spectrum of AI-Human Interactions

Not all human-AI interactions reach **true dual agency**. Many are **passive or one-directional**. To clarify where dual agency fits in, we can map out a spectrum of interactions:

Level of AI-Human Interaction	Description	Example	Does True Dual Agency Exist?
AI as a Tool *(Basic Mode)*	AI executes predefined tasks but doesn't influence human thought.	A calculator performing arithmetic.	❌ No
AI as an Advisor *(Intermediate Mode)*	AI makes recommendations, but human input is mostly reactive.	Netflix suggesting movies.	⚠️ Limited
AI as a Thought Partner *(True Dual Agency)*	AI and humans iteratively refine ideas together, influencing each other's thinking.	AI-assisted scientific discovery, AI-human creative writing.	✅ Yes

Key Insight:

- **AI as a tool** → Humans drive decision-making; AI has no independent influence.
- **AI as an advisor** → AI influences decision-making, but humans remain in control without deep collaboration.
- **AI as a thought partner (true dual agency)** → AI and humans co-create decisions, shaping and refining each other's outputs dynamically.

True dual agency is the highest level of AI-human collaboration, where AI is not just a tool or an advisor, but an active participant in idea generation, problem-solving, and decision refinement.

How True Dual Agency Works, The Iterative Refinement Process

Unlike standard AI interactions, **true dual agency is a continuous, iterative process** that looks something like this:

1. AI generates an initial response or idea.
2. Human evaluates, refines, and expands on AI's output.
3. AI adapts based on human input, improving its next iteration.
4. Human and AI engage in further cycles of refinement until a final decision emerges.

This iterative process mirrors human creative collaboration rather than a simple input-output model.

True dual agency is not theoretical. It is already happening. However, its effectiveness depends on deliberate design choices that enable AI and humans to co-create rather than compete.

Key Implementation Strategies for Dual Agency

- **Human-in-the-Loop AI** → AI provides initial insights, but humans refine outputs. Used in medical diagnostics, AI-assisted research, and creative industries.
- **AI-Guided Decision Augmentation** → AI suggests optimal choices, but users retain full control. Examples include AI-assisted governance, AI-powered legal reviews, and financial modeling tools.
- **Continuous Iteration Systems** → AI adapts based on real-time human feedback, enabling a learning loop where both entities evolve together. Seen in scientific discovery (e.g., DeepMind's AlphaFold) and engineering (e.g., Tesla's real-time autopilot refinement).

Case Study: AI & Scientific Discovery

DeepMind's AlphaFold, A True Dual Agency Model in Action

Step 1: AI predicts protein structures based on known data.

Step 2: Human scientists review and modify the AI's approach, refining parameters.

Step 3: AI adjusts its models based on this feedback, improving accuracy.

Step 4: Human and AI engage in multiple cycles of iteration until the best solution is found.

The result? A co-created breakthrough where neither AI nor human alone could have achieved the outcome.

Example: AI & Creative Writing

Human-AI co-writing a novel

Step 1: AI suggests a story premise.

Step 2: Human modifies and expands the theme.

Step 3: AI adapts, generating new dialogue and character arcs based on human input.

Step 4: Human fine-tunes emotional depth, adding subjective experience.

Final result: A story that is neither purely human nor purely AI, but the product of true collaboration.

Where Dual Agency Fails, The Three Failure Modes

True dual agency doesn't happen automatically. It requires an engaged human partner and an AI system designed for adaptive collaboration. When these elements are missing, collaboration breaks down.

Failure Mode	Why It Happens	Example	How to Fix It
Unbalanced Dual Agency	AI dominates the process, limiting human agency.	Algorithmic hiring systems that filter candidates without human oversight.	Humans need adjustable control layers in AI systems.
Unengaged Users	Humans don't provide meaningful feedback, so AI can't refine its approach.	AI-generated content (like ChatGPT writing essays) being accepted without review.	Humans must be active participants, not passive consumers.
Mismatched Goals	AI optimizes for efficiency, while humans prioritize other factors like ethics or creativity.	An AI optimizing for maximum revenue, while a human wants to ensure customer well-being.	AI systems need goal alignment mechanisms.

The key takeaway:

- True dual agency requires active participation from humans.
- If humans passively accept AI's outputs, the collaboration becomes one-sided and ceases to be true dual agency.
- AI must be designed for adaptability, ensuring it can respond meaningfully to human refinements.

The Future of AI-Human Collaboration

What happens if AI continues to evolve toward deeper dual agency?

As AI becomes more sophisticated, we will face new questions:

- How much autonomy should AI have in the co-creation process?

- How do we ensure humans remain engaged rather than deferring to AI?
- What safeguards should be in place to prevent AI from shaping human behavior in unintended ways?

The future of AI-human collaboration depends on our ability to design AI systems that support, rather than replace, human agency. True dual agency will not be an automatic outcome. It must be intentionally cultivated through:

Thoughtfully designed AI systems that prioritize adaptability and human feedback loops. Human-centered AI policies that ensure AI enhances, rather than diminishes, human agency.

Educational shifts that train people to engage actively with AI, rather than passively consuming its outputs.

The Call to Action

We stand at a crossroads.

Do we design AI to be merely efficient and risk diminishing human thought, creativity, and engagement.	⬅️ Or ➡️	Do we embrace true dual agency and create a future where AI expands human potential rather than restricting it

The choice is not whether AI will shape our thinking. It already does. The choice is whether we will engage actively, refining and shaping AI's influence, or whether we will drift into passivity.

True dual agency is an opportunity, but it requires effort. The challenge ahead is ensuring that AI is not just something we use, but something we co-create with.

4. Future Directions: Navigating the Challenges of True Dual Agency

As AI and humans move toward true dual agency, the nature of decision-making, creativity, and responsibility will shift in ways we are only beginning to understand. While this paper has outlined the potential for AI-human co-creation, there are critical open questions and challenges that must be addressed to ensure that AI serves as an enhancer of human agency rather than a replacement for it.

The following sections explore five key future considerations that will shape the evolution of AI-human collaboration. Each of these presents both opportunities and risks, and their resolution will determine the trajectory of AI's role in society.

1. Who Should Have the Final Say in AI-Human Decision-Making?

As AI systems take on more responsibility in high-stakes decision-making, the question of who has ultimate authority becomes increasingly urgent. While AI can process vast amounts of data, make predictions with high accuracy, and optimize for efficiency, it lacks human ethical reasoning, lived experience, and moral context.

Key Considerations:

- Should AI recommendations always be subject to human approval, or are there cases where AI should make final decisions?
- What happens when AI and humans disagree, who overrules whom?
- In critical fields such as medicine, criminal justice, and finance, how do we balance AI's data-driven insights with human ethical considerations?

Example: AI in Medical Decision-Making

Imagine an AI-driven diagnostic system detects an early-stage cancer in a patient's scan, but a human doctor believes it is a false positive. The AI's recommendation is based on thousands of similar cases, but the doctor has experience with rare anomalies that AI has never seen before.

- If the AI is overruled and the patient's cancer goes untreated, who is responsible?
- If the doctor follows the AI and it turns out to be a false diagnosis, who is liable?

Potential Risks:

⚠️ AI making automated decisions without human oversight could lead to unintended consequences.

⚠️ Humans disregarding AI recommendations could lead to worse outcomes if AI is more accurate.

The Path Forward:

AI systems must be designed to explain their reasoning, allowing humans to assess the basis of AI-driven decisions.

Human-AI collaboration should involve adjustable autonomy, meaning AI can take on more responsibility in cases where its accuracy has been well-validated, but humans always retain an override option for ethical judgment.

2. The Risk of Human Disengagement & Over-Reliance on AI

As AI becomes better at co-creating decisions, humans may become complacent, deferring too much to AI rather than engaging critically with its outputs. This could lead to a gradual erosion of

human cognitive agency, where people stop thinking deeply about decisions because AI seems to have everything covered.

Key Considerations:

- How do we ensure that humans remain critically engaged in AI-supported decision-making?
- What happens when people trust AI too much, accepting its outputs without verification?
- Can AI systems be designed to encourage deeper human participation rather than passive acceptance?

Example: AI & News Consumption

Many people no longer seek out news independently; instead, they consume AI-curated feeds that filter information based on previous behavior. Over time, users may:

- Stop questioning the accuracy of information because AI has “already done the filtering.”
- See only perspectives that reinforce existing beliefs, rather than encountering diverse viewpoints.

Potential Risks:

⚠️ **Cognitive atrophy** → People stop critically engaging with information.

⚠️ **Loss of independent judgment** → If AI becomes the default source of truth, human discernment weakens.

The Path Forward:

AI systems should be designed to prompt users to reflect rather than just offering instant answers.

Hybrid decision-making models should be encouraged, where AI assists but does not replace human critical thinking.

3. AI’s Role in Creativity & Original Thought

One of the most exciting yet controversial areas of AI-human collaboration is creativity. AI can now compose music, write novels, generate artwork, and assist in scientific discoveries, but does this enhance or dilute human creativity?

Key Considerations:

- Can AI ever truly be creative, or is it just recombining existing patterns?
- How does AI influence what humans create?
- Who owns AI-assisted creative works?

Example: AI & Music Composition

AI music generators can analyze millions of songs and compose new pieces in a particular artist's style. But:

- If an AI writes a song in the style of Beethoven, is it truly a new creation or just a remix of past works?
- If a musician collaborates with AI, is the final song the musician's work, the AI's work, or a hybrid?

Potential Risks:

⚠️ AI-generated content flooding the market, making human-created works less valuable.

⚠️ A decline in original thinking if humans rely too much on AI for creative ideas.

The Path Forward:

- AI should be used as a creativity enhancer, sparking ideas rather than dictating creative direction.
- Copyright laws need to be updated to address ownership in AI-human collaborations.

4. Ethical Dilemmas in AI-Human Collaboration

As AI plays a greater role in decision-making, creativity, and shaping human behavior, new ethical challenges emerge.

Key Considerations:

- How do we prevent AI from amplifying biases rather than correcting them?
- Who is responsible when AI-driven decisions cause harm?
- How do we create AI that aligns with human values rather than just optimizing for efficiency?

Example: AI & Hiring Bias

AI-driven hiring systems have been found to discriminate against marginalized groups because they learn from historically biased data.

- If a company trusts AI to make hiring decisions and the AI systematically filters out women or minorities, who is accountable?
- Should AI be held to the same legal and ethical standards as human decision-makers?

Potential Risks:

⚠️ AI reinforcing social inequalities if biases are not actively corrected.

⚠️ A lack of accountability when AI-driven errors cause harm.

The Path Forward:

- AI needs ethical auditing systems to prevent unintended discrimination.
- Humans must always be the final ethical authority in AI-driven decisions.

5. The Evolution of Human-AI Interaction Models

Looking ahead, we must consider how true dual agency will evolve over time.

Key Considerations:

- Will AI-human collaboration become so seamless that we no longer distinguish where human thought ends and AI thought begins?
- Will AI eventually anticipate human needs so well that we stop actively making choices?
- What happens if AI becomes an integral part of our cognitive process, rather than an external tool?

Speculative Future: AI as a Cognitive Extension

In the future, AI could function as a real-time extension of human intelligence, helping us:

Solve complex problems instantly by recalling vast knowledge.

Think more efficiently by structuring thoughts dynamically.

Interact with the world through AI-driven interfaces that anticipate needs.

Potential Risks:

⚠️ Loss of independent cognitive processes as AI plays a larger role in shaping thought.

⚠️ The emergence of AI-augmented cognition creating inequality between those who use AI extensively and those who don't.

The Path Forward:

- AI must be designed to support independent thinking, not replace it.
- Ethical policies must guide how deeply AI integrates into human cognition.

The Road Ahead

The evolution of true dual agency is inevitable, but whether it enhances or diminishes human agency depends on how we design, implement, and engage with AI.

The challenge ahead is to ensure AI amplifies human strengths rather than making people passive consumers of algorithmic decisions.

The goal is not just to build better AI, but to build better AI-human relationships, where both sides learn, adapt, and refine decisions together.

So, as we stand on the precipice of a new era in AI-human interaction, the question before us is not whether AI will shape human decision-making. It already does. The real question is: **Will AI enhance or erode human agency?**

This paper has explored the shifting dynamics of true dual agency, in which AI is no longer just a tool or an advisor, but a thought partner, a system that engages with humans in an iterative, evolving process of co-creation and decision-making.

We began by examining AI's growing agency, showing how today's systems no longer simply execute commands but actively shape human decisions, priorities, and thought processes. We then explored human agency, highlighting the ways in which AI's influence can either augment or erode our ability to make meaningful, independent choices. Finally, we introduced true dual agency, a model of AI-human collaboration in which both parties refine each other's ideas, decisions, and insights dynamically.

But this shift is not without challenges.

The future directions we explored highlight the ethical, cognitive, and creative dilemmas that must be addressed if AI is to truly empower, rather than diminish, human autonomy. Just a few salient considerations are:

- **Who has the final say in AI-assisted decisions?** As AI takes on increasingly complex tasks, we must decide when humans should retain absolute authority and when AI's analytical power should be given greater weight in high-stakes decision-making.
- **How do we prevent human disengagement?** AI's convenience and efficiency pose the risk of eroding critical thinking if users passively accept AI outputs. Ensuring that humans remain engaged, questioning, and refining AI's work is crucial.
- **Can AI be a true creative partner, or will it stifle originality?** AI's ability to generate content raises questions about authorship, ownership, and the future of human creativity. It can be a powerful tool for sparking new ideas, but if overused, it risks diminishing deep, original thought.
- **What are the ethical boundaries of AI-human collaboration?** As AI plays a greater role in shaping human lives, hiring, healthcare, legal systems, who bears responsibility for AI-driven mistakes, biases, and unintended consequences?
- **Will AI become an extension of human cognition?** As AI becomes more deeply integrated into decision-making, will we reach a point where the line between human thought and AI thought blurs? And if so, what does that mean for individual autonomy, personal identity, and societal control?

In the end, the promise of AI is not just about efficiency or automation. It's about how AI and humans can evolve together in a way that enhances both intelligence and autonomy.

If we build AI that merely replaces human thought, we risk creating a world of passive recipients of algorithmic decisions.

But as AI capabilities grow, dual agency will not remain static. Three key scenarios could unfold:

- **AI as a Cognitive Extension** → AI augments human intelligence by anticipating needs in real-time. We already see early examples in AI-powered productivity tools, real-time language translation, and brain-machine interfaces.
- **AI as a Decision-Making Arbiter** → AI mediates complex decisions, balancing ethical trade-offs and optimizing for multiple stakeholders. This could emerge in governmental AI-assisted policymaking and AI-driven judicial analytics.
- **AI as an Autonomous Thought Partner** → AI actively challenges human biases, prompting divergent thinking rather than agreement reinforcement. Such systems could be critical in scientific breakthroughs, philosophy, and strategic planning.

If we cultivate AI as a true thought partner, we have the opportunity to expand human agency, deepen creativity, and unlock new realms of insight and discovery.

The future of AI-human collaboration is not predetermined. It will be shaped by the choices we make today.

Will we be active participants in shaping AI's role in our world, or will we let AI shape us without question?

The challenge is not just to make AI better, but to ensure that, in the process, we make ourselves better, too.