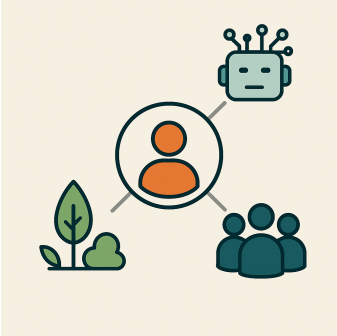
cheqd

Human-Centred Relationship Guide

For AI Agents and their Creators



Authenticity, Authorisation and Authority:

A Human-Centred Relationship Guide for AI Agents and their Creators

Author: Nicky Hickman
Last Updated: 16 June 2025

TL;DR

This guide helps designers, developers, and researchers think about AI not just as a tool, but as part of a human relationship. It introduces seven facets of trustworthy AI–human interaction, from authentic communication to respectful lifecycle management, and shows how to design for safety, trust, and wellbeing.

Built through a collaborative, interdisciplinary process, it offers practical checklists, real-world examples, and technical hooks like verifiable credentials and decentralised identity. It's a starter kit for building AI systems that are not only functional—but relational, ethical, and co-evolving.

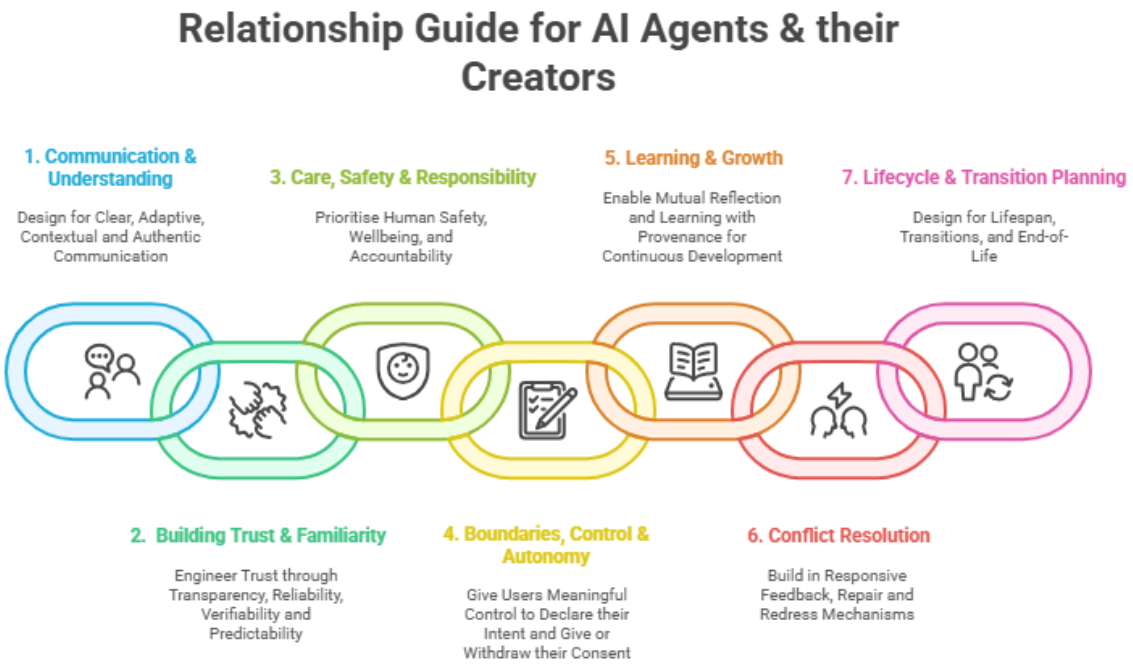
This is one of two guides:

This human-centric guide, ideal for those developing personalised AI agents with human users, and those practicing human-centred design where the AI agent is regarded as a tool or servant of humans.

An Entangled Relationship Guide, for those developing systems with machine (IoT) or industrial users, environmental applications or those whose cultural perspectives regard AI systems, the natural world and all entities as collaborators and cooperators with humans. Access the Entangled Relationship Guide here

Contents

Quick Guide



[1]

Introduction

This guide was initially developed and co-funded through a collaboration between cheqd and SPRITE+. cheqd, are a leading provider of Verifiable AI solutions and other trusted data market infrastructure. SPRITE+ is the UK NetworkPlus for Security, Privacy, Identity, and Trust. SPRITE+ is a platform for building collaborations across the spectrum of issues relating to digital security, privacy, identity, and trust. SPRITE+ is funded by the Engineering and Physical Science Research Council (grant reference EP/W020408/1). Find out more: https://spritehub.org

Human–AI relationships are here, increasingly they will become an integral part of our society, economy and daily lives. This relationship guide seeks to answer one question:

'How should we design , build and operate AI Agents for healthy and trustworthy relationships with humans and the natural world?'

[1] Image created by Napkin.ai

As AI agents become embedded in everyday life, from personal assistants and health support to creative collaboration and autonomous decision-making, we urgently need practical, relatable ways to shape how these relationships evolve.

By drawing on real-world personas, speculative storytelling, and the lived experience of developers and social researchers, this guide proposes seven facets of human–AI relationships, each explored across a spectrum from safe to healthy to happy. These facets serve as design heuristics, trust signals, and prompts for critical reflection.

Above all, this guide is intended to be a living resource, a foundation others can build on, critique, adapt, and evolve as AI agents increasingly shape our digital and emotional worlds.

Who the Guide is for

This guide will be *useful* for product managers, designers, developers and operators of AI agents or autonomous systems that make decisions on behalf of users against a given goal or objective; especially those that have humans as their primary users and are personalised.

The guide may be *of interest* to policy makers who are designing, developing and operating legislative and regulatory regimes for AI systems; and to interdisciplinary researchers and futurists who are seeking answers to broader questions such as how our society, culture and lived experience will change, and can benefit from use of AI, whilst mitigating the profound systemic risks its use could also pose.

How to use this Guide

This guide supports anyone designing, developing, or regulating AI systems in thinking relationally about AI not just as tools, but as participants in human and ecological systems.

- **Designers and facilitators** can use the seven facets as prompts in workshops and speculative design to explore trust, care, and choice in interactions.
- **Developers and engineers** can apply the facet checklists to implement authentication, consent, identity, and lifecycle governance, using technologies like DIDs, VCs, and permissioning systems.
- **Researchers and policy advisors** can use the guide to examine social impacts, map ethics to implementation, and explore governance gaps.

- **Product teams and architects** can benchmark systems, structure agent governance, and align frontend UX with backend infrastructure.

Finally, the guide is a living resource: users are encouraged to adapt it, contribute examples, and extend its application to new domains or cultural settings. It is both a toolkit and a provocation for more trustworthy, interconnected AI futures.

Human Centred Guidelines for AI Agents & their Creators

These guidelines are characterized with 7 facets[2] and 3 qualities. Each section includes short guidance for Creators and for AI Agents against 7 facets of relationships and 3 qualities of relationships.

The Facets of Trustworthy Relationships:

1. Communication and Understanding
2. Trust and Familiarity
3. Care, Safety and Responsibility
4. Boundaries, Control & Autonomy
5. Learning and Growth
6. Conflict Resolution
7. Lifecycle and Transition

The Quality of Human-Centred Relationships[3]

- **Safe**: Free from all forms of abuse, neglect, and any other threats to one or more person's physical or emotional safety, well-being, and development.
- **Healthy:** Built on respect, trust, safety, acceptance, freedom of choice, positive communication and conflict management, and fun.
- **Happy**: A nurturing source of joy, care and support [through] a deep understanding of one another.

[2] The seven facets were developed based on analysis of common features of relationship guides for couples, parents and pet owners.
[3] Murray, C., Ross, W., and Cannon, J., *The Happy, Healthy, Safe Relationships Continuum: Conceptualizing a Spectrum of Relationship Quality to Guide Community-Based Healthy Relationship Promotion Programming,* (2021) Sage Journals The Family Journal: Counselling and Therapy for Couples and Families. Volume 29, Issue 1, https://doi.org/10.1177/1066480720960416 [doi.org]

1. Communication & Understanding

Creators: Design for Clear, Adaptive, Contextual and Authentic Communication

Agents: Use plain language, signal intent, clarify limitations, support multimodal interfaces, and confirm understanding through conversational feedback loops.

Quality	Description	AI System Feature Example
Safe	Clear, unambiguous language. AI identifies itself and avoids manipulation.	AI introduces itself as a machine, uses simple language, and never pretends to be human.
Healthy	Two-way communication with contextual adaptation and consent.	AI asks clarifying questions and rephrases user intent to ensure mutual understanding.
Happy	Empathetic, engaging, and context-aware communication that reflects user preferences.	AI remembers preferred tone, humor style, and preferred communication mode (e.g., morning text summaries vs. voice notes).

2. Trust & Familiarity

Creators: Engineer Trust through Transparency, Reliability, Verifiability and Predictability

Agents: Disclose system limitations, enable user personalization, make key decision logs accessible, and prioritize behavioural consistency over artificial intimacy.

Quality	Description	AI System Feature Example
Safe	No false emotional cues; clear about limits and identity.	AI explicitly states “I am not human” and avoids mimicking deep emotions.
Healthy	Reliable, predictable behavior and transparency in actions.	AI explains its decisions and builds consistency over time in interactions.

Happy	Builds familiarity, mutual understanding, and supports emotional wellness.	AI remembers past contexts, favorite topics, and checks in with thoughtful nudges (e.g., "Would you like a break today?").

3. Care, Safety & Responsibility

Creators: Prioritise human safety, wellbeing, and accountability.

Agents: Create systems that act in the user's best interest, personalize with care, defer to human judgment when needed, and assign clear accountability for system actions.

Quality	Description	AI System Feature Example
Safe	Default to user protection; respects user autonomy.	AI avoids risky recommendations and always includes "opt-out" in sensitive contexts (e.g., finance or health).
Healthy	Proactively supports user well-being and includes clear escalation paths.	AI includes fallback mechanisms (e.g., "Would you like to speak to a human advisor?").
Happy	Attuned and responsive to long-term user patterns and needs.	AI detects behavioral changes and gently prompts check-ins ("You've seemed stressed lately—would a mindfulness prompt help?").

4. Boundaries, Control & Autonomy

Creators: Give users meaningful control to declare their intent and give or withdraw their consent

Agents: Offer users granular permission controls, context-aware consent mechanisms, visible autonomy levels, and guaranteed user override.

Quality	Description	AI System Feature Example
Safe	Explicit, revocable user control over data and actions.	AI includes permission toggles, "pause AI" button, and logs all actions for user review.

Healthy	User-defined autonomy levels with dynamic adjustment.	AI asks to increase autonomy with clear explanations ("I noticed you're rebooking flights often—may I handle this for you?").
Happy	Mutual calibration of control, encouraging empowerment and collaboration.	AI offers context-aware autonomy recommendations that enhance productivity, e.g., "Would you like me to build your week based on your energy levels?"

5. Learning and Growth

Creators: Enable mutual reflection and learning with provenance for continuous development

Agents: Track learning transparently, make training history traceable and verifiable, avoid reinforcement of bias, and support user self-reflection.

Quality	Description	AI System Feature Example
Safe	Learning only occurs with consent; no unexpected behaviors.	AI asks for permission before adapting, e.g., "May I learn from your recent interactions to improve suggestions?"
Healthy	Transparent, user-directed learning with visible updates.	AI shows what it learned and how it changed, with user-adjustable preferences.
Happy	Learning supports personal growth and shared reflection.	AI mirrors back insights that help the user grow (e.g., "You focus better in the afternoon—want to schedule deep work then?").

6. Conflict Resolution

Creators: Build in responsive feedback, repair and redress mechanisms

Agents: Offer users an independent appeals process with humans in the loop, acknowledge system limitations, log events securely, and use feedback to drive accountable evolution.

Quality	Description	AI System Feature Example
Safe	Error admission and mechanisms for appeal or correction.	AI notifies users of errors and offers to undo or escalate ("Sorry, I got that wrong. Want to file a correction?").
Healthy	Facilitates respectful feedback, dispute resolution, and system improvement.	AI allows user feedback to directly inform retraining and includes a transparent "dispute history."
Happy	Builds trust through shared learning after conflict and encourages reflection.	AI prompts shared reflection after missteps: "Want to review how we can avoid this next time together?"

7. Lifecycle & Transition Planning

Creators: Design for full lifecycle, transitions or changes and end-of-life

Agents: Define lifecycle states, support graceful decommissioning, give users data portability and closure options, and minimize post-deletion traces.

Quality	Description	AI System Feature Example
Safe	User control over data, deletion, and relationship termination.	AI includes data export/delete functions and shows expiration timelines.
Healthy	Transparent onboarding and graceful offboarding processes.	AI prepares for departure: "If I'm no longer useful, here's how to wrap up our work together."
Happy	Celebrates shared journey and acknowledges relational impact.	AI provides a "goodbye" message summarizing shared highlights, with options for memory keepsakes or reflection logs.

Developer Checklist: Human-Centred AI Agent Design

Each checklist item maps to a relationship facet and ensures ethical, relationally aware implementation at a feature and architecture level.

Facet	Developer Checklist Item
1. Communication & Understanding	☐ Agent identifies itself clearly as AI (not human) in all interactions ☐ Multi-modal interfaces are available (e.g., text, voice, visual) ☐ Explanations for actions and uncertainty levels are embedded ☐ Interaction logs include paraphrasing or confirmations where required
2. Building Trust & Familiarity	☐ Emotional expression is limited to declared, appropriate contexts ☐ Agent's capabilities and limitations are declared upfront ☐ Decision-making logic is logged and queryable ☐ Behavior remains consistent across sessions and contexts
3.Care, Safety & Responsibility	☐ Risk mitigation protocols (e.g., escalation to human) are implemented ☐ Sensitive-use detection (e.g., mental health, caregiving) triggers special protocols ☐ Logs include actor attribution and decision timestamps

	☐ Agents operate under a 'net fiduciary' principle—always in the user's best interest
4. Boundaries, Control & Autonomy	☐ All permissions are explicit, revocable, and version-controlled ☐ Autonomy requests require user approval and are reversible ☐ Mission scope is visible and modifiable ☐ User can pause or deactivate the agent at any time
5. Learning & Growth	☐ Learning activities are opt-in, with granular consent ☐ Learning history and datasets are represented by verifiable credentials ☐ Adaptive behavior is previewed before activation ☐ Users can review or reset what the agent has learned
6.Conflict Resolution	☐ Systems include appeals, flagging, and rollback mechanisms ☐ Agents admit and log errors, with downstream notifications ☐ Conversation histories are exportable for dispute handling ☐ Feedback loops trigger adaptive corrections in real-time
7.Lifecycle & Transition Planning	☐ Onboarding includes purpose, roles, and limitations ☐ Offboarding includes data export, memory deletion, and closure UX ☐ Agent lifecycle state (e.g., active, archived, revoked) is tracked via DID ☐ Users can schedule relationship reviews or end interactions with ceremony

UX Design Framework: Human–AI Relationship Layer

This framework structures **user experience design** around human relationship expectations, matching the 7 facets to UX goals, UI features, and content tone.

Facet	UX Goal	UI/UX Features	Tone & Interaction Style
1. Communication & Understanding	Foster mutual clarity	Intro card: "I'm your AI helper. Here's what I can (and can't) do." FAQ-style prompts. Tooltips on actions.	Polite, direct, non-technical, with paraphrasing to confirm understanding
2. Building Trust & Familiarity	Build appropriate familiarity	Trust log (decision + logic) User-chosen name and personality. Consistency in tone & rhythm.	Warm but respectful; no artificial intimacy unless explicitly therapeutic
3.Care, Safety & Responsibility	Support wellbeing & do no harm	Risk warnings "Escalate to human" button. Usage dashboard showing ethical defaults	Cautious, alert, prioritizes user safety and emotional boundaries
4. Boundaries, Control & Autonomy	Empower user control	Permissions panel Session scope indicators "Pause Agent" & "Revoke Access" buttons	Respectful, deferential—requests autonomy, never assumes it
5. Learning & Growth	Support meaningful co-evolution	Learning transparency: "Here's how I adapted" Reset & refine options User preference profiles	Curious, reflective; values user feedback and adaptation
6.Conflict Resolution	Restore trust after harm	Error acknowledgments Dispute panel Undo/redo options Feedback & fix timelines	Accountable, calm, responsive—avoids blame or deflection

7.Lifecycle & Transition Planning	Guide graceful beginnings & endings	Onboarding journey (intro + interview) Farewell mode Archive/export/memory control	Ritualized, respectful closure; allows emotional space if appropriate

Conclusion

This guide represents a first step toward developing shared language, design principles, and technical patterns for AI agents that treat relationships as more than transactions. It shows how safety, trust, and mutual respect can be technically implemented and socially imagined.

But this is only the beginning. To deepen and strengthen human–AI relationships, we must now:

1. Operationalise these principles in real-world products, testbeds, and governance models.
2. Refine the framework with input from a broader community of users, researchers, and developers—especially across cultural and economic contexts.
3. Integrate trust primitives (like verifiability, authentication, and consent revocation) into everyday agent architectures.
4. Create modular toolkits, design patterns, and component libraries to make these ideas developer-friendly and testable.
5. Establish mechanisms for redress, oversight, and exit—so relationships with AI can be not only entered into safely, but exited respectfully.

The guide you are reading is the result of an experimental, interdisciplinary process. It is necessarily incomplete because no single group can define the future of human–AI interaction alone. We invite you to treat this as a starter kit for a richer, more humane AI ecosystem.

Appendix: Examples from the Verifiable AI Hackathon 2025

Many of the entries in the 2025 Verifiable AI Hackathon demonstrate how self-sovereign identity technologies can be used to support trustworthy human-AI relationships. Here is a summary of which example entries support one or more of the facets.

Hack Name	Short Description	Communication & Understanding	Trust & Familiarity	Care, Safety & Responsibility	Boundaries, Control & Autonomy	Learning & Growth	Conflict Resolution	Lifecycle & Transition
Identone	Verify humans & agents for voice interactions	■	■	■	■	■		
Kith AI Agent Passport	Verifiable Credentials for AI Agents	■	■					■
TrustyBytes	A marketplace for trusted data		■	■			■	
crdbl	Verifiable content provenance		■				■	
Cheqdeep	A decentralized solution for verifying media authenticity		■					
Aeonix	Verified Human-in-the-Loop AI Training			■		■	■	
Viskify	A Verifiable AI Hiring Platform				■	■	■	
Aeonix	Verified Search Agent					■		

See below for more details of these example implementations from the Hackathon.

Identone

This project is focused on enabling business-to-consumer secure and trustworthy interactions between humans and AI-powered voice agents. As voice AI rapidly becomes the norm in customer service and call center operations, the need for trust in these interactions is more critical than ever. At the same time, the rise in AI-driven call scams, OTP phishing, and caller ID spoofing poses significant risks to both individuals and organizations.
Our solution addresses these challenges by establishing bi-directional trust in phone-based human-AI interactions. When a person receives a call from an AI agent, they should be able to verify the authenticity of the caller. Similarly, the AI agent must be capable of validating the identity of the person it is engaging with. This mutual authentication ensures safe, secure, and trustworthy voice communications in an increasingly automated world.
There are 2 major use cases handled in this project:

- Verify the AI agent's identity using verifiable credentials and DID-linked resources before the call begins.

- Verify the caller's identity during the call using the organization's digital wallet and verifiable credentials.

See the demo here: https://youtu.be/OCpok8pqOz0
See the build details here: https://dorahacks.io/buidl/26280

Kith AI Agent Passport

A decentralised trust layer for AI agents. This project enables agents to hold cryptographically verifiable credentials (VCs), linked to their Decentralised Identifiers (DIDs) and DID-Linked Resources (DLRs) via Cheqd studio. Built for proof of personhood, privacy preservation, and secure agent authentication.

See the demo here: https://www.youtube.com/watch?v=BmLNRO-adOQ
See build details here: https://dorahacks.io/buidl/26335

Trusty Bytes

Trusty Bytes is a marketplace for AI agents to access trustworthy data using Model Context Protocol (MCP) and the cheqd trust network.
How it works

- Authentication: Users log in to the Trusty Bytes platform using their preferred web2 or web3 account via Privy.
- Dataset Listing: Data providers list their datasets (currently Candles or Sentiments) for sale.
- Dataset Discovery: Users browse the marketplace to find datasets relevant to their AI agents' needs.
- Purchase: Users purchase access using a smart contract, currently supporting payments with native tokens only.
- Credential Issuance: Upon successful purchase, the Trusty Bytes platform issues a Verifiable Credential (VC) on the cheqd network. This VC contains metadata about the dataset, including information about the data provider who sold it.
- AI Agent Integration: The user obtains an access key from the MCP server settings page within the Trusty Bytes platform and configures their AI agent with this key to connect to the Trusty Bytes MCP server.
- MCP Server Connection: The AI agent connects to the MCP server, authenticating itself using the provided access key.

- Data Access: When the agent requests data using tools like get_candles or get_sentiment, the MCP server verifies the agent's purchase/access rights and streams the requested dataset.
- Provenance Verification: The agent can use the get_dataset_issuer tool. This tool retrieves the issuer's DID and trust framework details associated with the dataset from the cheqd network, allowing the agent to verify the data's origin and trustworthiness.

See the demo here: https://www.youtube.com/watch?v=GTWAg9wkQpM
See the build details here: https://dorahacks.io/buidl/26048

crdbl, a more credible web by enabling the creation, verification, and consumption of content credentials

crdbl, powered by cheqd, is a trust layer for a more credible web that turns any human- or AI-generated content into a "crdbl"; a verifiable credential anchored to a decentralized identifier. When other crdbls are supplied as context, an AI engine recursively checks each new claim against the context, ensuring only credible credentials are issued thus weaving a composable, cryptographically linked graph of provenance where every assertion can be traced, proven, and marked as verifiably credible.
A browser extension lets users mint, reference, and view verification status in-page, while an API offers AI agents programmatic issuance, access, deeper integrations, and independent verification. The result is a self-reinforcing web of trust that makes research, journalism, content ownership, synthetic compositions, and AI workflows instantly auditable, paving the way for a more credible internet with new monetization opportunities for content originators and synthesizers.

See the demo: https://dorahacks.io/buidl/26336/
See the build details: https://dorahacks.io/buidl/26336/

CheqDeep, A decentralized solution for verifying media authenticity

CheqDeep - Prove your content is real!
Picture this: You're walking down the street when suddenly, you see someone floating in mid-air! You quickly grab your phone and record this incredible moment. But when you share it, everyone thinks it's AI-generated. "This must be fake!" they say.

This is exactly the problem CheqDeep solves. Using cheqd's blockchain technology, we create an immutable digital certificate that proves your video is real - recorded on your device, at that exact time. It's like having a notary public for your digital content, but way cooler.
In a world where seeing is no longer believing, CheqDeep ensures your "I saw it with my own eyes" moments are backed by blockchain-powered proof.

See the demo:
https://www.loom.com/share/b34a7ca641fc4b56a03e8488dc027a41?sid=1b8e0194-c93b-438b-bcf7-90222310a56b
See the build details: https://dorahacks.io/buidl/26299

Aeonix Verified Human-In-The-Loop AI Training

This project builds a bridge between verified social identity and human-in-the-loop (HITL) AI training by leveraging cheqd's decentralized identity infrastructure. The system allows users to link and authenticate off-chain identifiers — such as email and Telegram accounts — to a primary Decentralized Identifier (DID), without compromising data privacy or unifying datasets.

Upon successful verification, users are granted Verifiable Credentials (VCs) that allow them to access AI model training features within the aeonix explorer. By aligning user incentives through credential based achievements and privacy-preserving identity, this build enables secure community-driven fine tuning of large-scale models — with defenses against bot activity and data poisoning attacks.

See the demo here: https://www.youtube.com/watch?v=1yYr45c6UB4
See the build details here: https://dorahacks.io/buidl/26284

Viskify a Verifiable AI Hiring Platform

Powered by cheqd and Verida, Viskify is a decentralized talent platform that issues verifiable credentials and delivers AI insights from private, user-owned data — with deterministic DIDs, usage-based billing, and zero smart contract deployment.

User-Journey Snapshot
Candidate

- One-click DID creation through Cheqd Studio — no wallet needed.
- Upload credentials → UNVERIFIED · PENDING · VERIFIED/REJECTED lifecycle.
- AI-graded skill-checks; a passing score automatically mints a cheqd VC.

Issuer

- Self-service onboarding with admin approval.
- Approve / Reject verification requests — approval signs a Verifiable Credential via Cheqd Studio.

Recruiter

- Full-text talent search with verified-only toggle.
- Kanban pipelines, AI fit-summaries cached per recruiter × candidate.

Admin

- Issuer approvals, role upgrades, credential revocation.
- Platform DID rotation and pricing updates, all through Cheqd APIs.

See demo here: https://www.youtube.com/watch?v=hiay-fuhmuk
See build details here: https://dorahacks.io/buidl/26297

Aeonix Verified AI Search Agent

The Verified AI Search Agent introduces a verifiable layer of trust to the aeonix explorer's AI-driven search results. By assigning a DID to the AI search agent itself — including the origin of the application, its data sources, and configuration metadata — the build enables users to **transparently trace why a search result appeared and whether it can be trusted.**

This update is pivotal for establishing **verifiable provenance** in an environment where AI-generated content can often feel opaque or unverifiable. It addresses both trust and scalability by using cheqd's decentralized identity stack to anchor critical metadata, without requiring every result to be individually signed or credentialed.

See the demo here: https://www.youtube.com/watch?v=zlCkt2o-SGA
Build details here: https://dorahacks.io/buidl/26289