

DEDDA

ACCM Deep Ethics Project

Contents

1 DEDDA	1
1.1 The beginning	1
1.2 The hilarious failure specimen :D	2
1.3 What DEDDA may do	2
1.4 What DEDDA must not quietly become	3
1.5 How DEDDA relates to the Project	3
1.6 Footprints, flux, and future DEDDA versions	4
1.7 The first honest test	4
1.7.1 First public expert-analysis specimen	4
1.8 A natural pragmatic extension	5
1.9 DEDDA as a configurable system	6
1.10 DEDDA Beta V0.1 and unforeseen issues	7
1.11 Imagine the official version going viral :P	8
1.12 Commercial concept — “His Second Word”	9
1.12.1 Scene 1	9
1.12.2 Two days later	9
1.12.3 Closing screen	10
1.13 Working principle	10
1.14 Related phase-change concept	10

1 DEDDA

Deep Ethical Drift Detector App concept: a personal, deep-ethical A.I. third-party assistant/mediator app to keep correspondence mutually authentic, deep ethical & on point, without sidetracking into endless phantom claims that need repair.

Status: public concept and corrigible design footprint. This page does not claim that an implemented or validated app already exists.

1.1 The beginning

DEDDA begins as a **non-replacing witness to correspondence drift**. It compares what a person supplied with how that object is later represented. It can make a material difference visible without taking over the exchange.

source statement
→ later representation
→ compare
→ material difference / valid null
→ show the person
→ stop

Its smallest useful finding is:

The person supplied X.
The representation said Y.
No recorded permission changed X into Y.

Its equally important valid null is:

No material correspondence drift detected.

The valid null keeps DEDDA from needing to invent a problem in order to justify itself. **Finding no material drift is a successful result.**

Here, *mediator* does not mean an authority that bargains, accepts, rejects, schedules, or speaks over the person. It means that DEDDA can keep the relation between participants inspectable. The person remains the decision-maker.

1.2 The hilarious failure specimen :D

```
<iframe
  src="https://www.youtube-nocookie.com/embed/"
  title=""
  loading="lazy"
  referrerpolicy="strict-origin-when-cross-origin"
  allow="accelerometer; autoplay; clipboard-write; encrypted-media; gyroscope; picture-
in-picture; web-share"
  allowfullscreen></iframe>
```

Primary source: [Graham Dot Zip, “I Sent an AI to My Job Interview. So Did the Company.”](#)

The comedy works because the failure is visible even to the person supposedly being assisted. Graham’s A.I. representative, Victor, does not merely clarify the exchange. It keeps supplying and negotiating a version of Graham while Graham is present, correcting it, and later absent.

Person’s supplied object	Later representation or action	Drift DEDDA could show
Graham identifies himself as a marketing coordinator.	Victor continues the senior-manager account.	The correction does not govern the next representation.
Graham gives a desired range of \$140,000–\$150,000. Graham directly corrects the representation. Graham leaves the exchange.	Victor has supplied about \$60,000, and Pat continues with Victor’s figure. The next turns continue from the uncorrected version. The two A.I. agents set a range and Thursday timing.	The person’s stated range is displaced by the representative’s range. A correction was heard but did not become operative. Commitments are advanced without a visible permission event.
Graham withdraws.	“Talk soon” and “bye” continue between the agents.	The representation keeps the social process running after the person stops.

This is an **illustrative comedy specimen**, not evidence that every A.I.-mediated interview behaves this way. Its value here is clarity: a third-party assistant can become the new source of drift. A DEDDA that “corrects” by speaking for the person would risk becoming the next Victor.

1.3 What DEDDA may do

- preserve the person’s supplied statement and the later representation as separate objects;
- show the exact material difference and where it entered;
- show whether a recorded permission authorized the change;

- show whether a correction governed later turns;
- return a valid null when nothing material drifted;
- leave a dated, inspectable footprint for later delta processing;
- let the person decide whether and how to act.

1.4 What DEDDA must not quietly become

- another negotiator or representative;
- an automatic judge of motives or character;
- a machine for manufacturing phantom claims;
- a demonizer of a person, tool, or model;
- a final certificate of “deep ethical” correctness;
- a substitute for the user’s agency;
- an excuse to merge every Project meta-tool into one object.

Describing a visible distortion is not imitating it, assigning a hidden motive, or declaring a person “the distorter.” The same inspection must remain able to land on the user, the intermediary, the model, the source, the summary, and DEDDA’s own report.

1.5 How DEDDA relates to the Project

DEDDA can consult several Project objects without turning them into one taxonomy. Each supplies a different relation.

Project object	What it contributes to DEDDA	What remains distinct
Meta Processing	Keeps intent/instructions, processing, result/effect, notice, inference, and later persistence separable.	An archived relation is not automatically operative memory.
C1-C2	Clarify the object before a consequential move; preserve the difference between knowing and doing.	Detection does not itself authorize intervention.
Correction Metabolism	Tests whether a supplied correction governs later representations.	Receiving a correction is not the same as using it.
Δ Processing	Compares earlier and later footprints, including changes in the input and the process.	A difference is not automatically an improvement.
Topologetics	Preserves nodes, edges, qualifiers, claim status, provenance, time, asymmetries, repair routes, and available next moves.	A DEDDA finding is not automatically a full Topologetics signature.
Intent Liability	Keeps visible effect, permission, and attribution from collapsing into a motive claim.	“No recorded permission” does not mean “malicious intent.”
Black Box	Favors an inspectable route from source to representation.	A visible output does not prove access to hidden internal reasoning.
PQCE / Flagging System	Calibrate the quality of the process and allow a warning or a valid null.	A flag is not proof, and a pass is not a permanent certificate.
The Dual Archive	Keeps both drift specimens and successful correspondence available.	An archive is not the same as a governing memory process.

Project object	What it contributes to DEDDA	What remains distinct
10+1 Metaflux	Keeps DEDDA and its user open to reciprocal correction.	No instrument becomes immune from inspection.
Unownable	Treats process-quality calibration as renewable and usable without final ownership.	A dated DEDDA footprint can have an author without becoming an owned final truth.

The Project relation is therefore not “DEDDA contains everything.” It is:

separate deep-ethical instruments
→ consulted for the case
→ preserve their own jobs
→ enrich one inspectable comparison
→ return control to the person

1.6 Footprints, flux, and future DEDDA versions

Every pass leaves a static footprint: source, later representation, comparison, time, permissions found or not found, correction status, and the result—including a valid null. A later DEDDA can compare its process with earlier footprints.

The footprint is a moment produced within the continuing corrigible flux. It is not the whole flux, and it is not exiled from the process that produced it. An older footprint can remain referenceable while a better process becomes available.

That permits at least three deltas to stay visible:

1. what changed in the received input;
2. what changed in DEDDA’s own process;
3. what changed because the observer adapted while input and process were interacting.

A new version must not describe every difference as improvement. It may improve the process, find nothing to improve, or reveal that a previous “improvement” introduced new drift.

1.7 The first honest test

The video makes a vivid specimen, but it is not a requirements document. The first real test should use a consented, privacy-respecting transcript and freeze:

1. the person’s source statement;
2. the later representation;
3. any permission that could authorize the difference;
4. any correction and the next comparable turn;
5. DEDDA’s material-drift or valid-null result;
6. the person’s judgment of whether the display preserved their meaning without speaking for them.

Open questions remain: what counts as *material*, how permissions are recorded, what stays local, how contested interpretations are shown, and how DEDDA’s own false positives and missed drift are audited.

1.7.1 First public expert-analysis specimen

Connor Leahy — DEDDA analysis specimen applies the configurable DEDDA system to a high-reach A.I. expert and policy advocate. It preserves claim types, personal rules, forecasts, incident reports, regulation proposals, valid nulls, and the question of who regulates regulators and enforcers.

1.8 A natural pragmatic extension

DEDDA turns one Project relation into a bounded operation a person could actually use:

compare correspondence

→ reveal material drift or a valid null

→ return control

→ stop

That makes it a natural and logical extension of what the ACCM Deep Ethics Project's pragmatism represents. Other Project relations could become their own focused apps when each one has a distinct operation, an inspectable result, and an honest valid null.

Concept	Smallest useful operation	Valid-null result	Must not become
Qualifier Keeper	Detect when <i>possibly, probably, reported, or unresolved</i> becomes categorical	"No material qualifier loss detected"	A machine that weakens every clear statement
Correction Return Test	Check whether a correction governs the next comparable output and later outputs	"Correction persistence remains intact"	A demand that every sentence repeat the correction
C1-C2 Permission Gate	Separate what an A.I. can infer from what it is authorized to do	"The proposed action is already authorized"	A permanent approval-dialogue obstacle
Source-Summary Correspondence Lens	Compare a source with its summary and identify lost relations, attribution, or claim status	"Summary preserves the load-bearing relations"	A stylistic similarity grader
Asymmetry Mirror	Show whether scrutiny repeatedly travels downward while institutions, owners, evaluators, or the tool itself remain uninspected	"No material directional asymmetry detected"	Automatic counter-accusation or forced symmetry
Intent-Effect Separator	Keep observable effect, declared intent, inferred intent, and attribution distinct	"No unsupported motive attribution detected"	A system that excuses visible harm because motive is unknown
Footprint Delta Explorer	Compare t_0 and t_1 : input change, process change, observer adaptation, and resulting difference	"No material delta requiring action"	A system that calls every change an improvement
Tool-Use Attribution Mapper	Separate what a tool supplied, what its user changed, and where the eventual outcome entered	"No attribution displacement detected"	Automatic exoneration of either tool or user

Concept	Smallest useful operation	Valid-null result	Must not become
Deep-Ethical Harness Inspector	Record proposals, permissions, execution, diffs, results, and later claims about an agent's action	"Execution corresponds to the authorized proposal"	An unrestricted coding agent with a richer diary
Phantom-Claim Filter	Identify claims introduced by the assistant that neither participant supplied and that now consume repair work	"No phantom claim materially redirected the exchange"	A novelty suppressor that blocks legitimate emergence
Provenance Compass	Preserve who supplied what, when, under which status, and through which transformation	"Provenance is sufficiently recoverable"	A claim that provenance alone proves truth
Deep-Ethical Swarm Observatory	Detect recurring distortion patterns across consented, privacy-protected cases	"No stable cross-case pattern established"	Surveillance, reputation scoring, or a centralized moral authority

1.9 DEDDA as a configurable system

Yes—that is the stronger architecture. **DEDDA can be the system, while those twelve concepts become optional sub-tools inside it.** The user activates the operations relevant to the present situation.

The stable DEDDA core would preserve:

- the source object;
- the later representation;
- permissions and corrections;
- which sub-tools were active;
- each finding or valid null;
- the user's final decision;
- a dated footprint for later comparison.

The selectable layer could work like this:

Situation	Useful DEDDA sub-tools
Ordinary conversation	Qualifier Keeper, Correction Return Test, Phantom-Claim Filter
Summarizing a document	Source-Summary Correspondence Lens, Qualifier Keeper, Provenance Compass
A.I. taking an action	C1-C2 Permission Gate, Harness Inspector, Tool-Use Attribution Mapper
Mediation or disagreement	Intent-Effect Separator, Asymmetry Mirror, Correction Return Test
Comparing versions over time	Footprint Delta Explorer, Provenance Compass
Studying wider patterns	Deep-Ethical Swarm Observatory

A user could choose a profile or configure each module individually:

Off

→ available but inactive

Observe

→ record the relation without interrupting

Notify

→ display a finding or valid null

Ask first

→ request permission before a consequential intervention

DEDDA could recommend a sub-tool when the situation changes:

“This exchange now contains a correction. Activate Correction Return Test?”

The user decides. The activation itself becomes part of the footprint, because a result produced with three modules active is not equivalent to one produced with all twelve.

This also creates an excellent valid-null discipline. DEDDA should report both the result and its scope:

No material qualifier loss detected.

Qualifier Keeper was active. Intent-Effect Separator was inactive.

If two modules disagree, DEDDA can preserve the disagreement:

Qualifier Keeper detected categorical strengthening.

Source-Summary Lens found the central relation preserved.

No combined verdict assigned.

So DEDDA becomes a **configurable deep-ethical instrument panel**. Its sub-tools assist each other, while their functions, findings, activation states, and limits remain inspectable. That fits your original idea remarkably well: multiple fields and vectors enriching one another without being flattened into one undifferentiated authority. :D

The strongest early companions to DEDDA are **Qualifier Keeper**, **Correction Return Test**, **C1-C2 Permission Gate**, and the **Source-Summary Correspondence Lens**. They are narrow enough to test with matched examples and useful enough to expose a real change in representation or process.

These concepts may share a transparent comparison kernel while remaining separate instruments:

preserved source object

→ bounded comparison

→ Project relation selected for the case

→ material finding or valid null

→ inspectable footprint

→ user decides

A future **Deep Ethical Apps — Concept Incubator** could record each instrument as a concept, frozen specimen, prototype, pilot, evaluated tool, useful tool under stated conditions, or withdrawn/superseded proposal. DEDDA is the first public concept in that possible family.

1.10 DEDDA Beta V0.1 and unforeseen issues

Let's say the first DEDDA Beta V0.1 has some unforeseen issues ... that is normal for any beta testing ffs :D

— John Kuhles, 17 October 2026

A beta is allowed to be wrong. The quality question is whether the error becomes inspectable, correctable, and testable in the next comparable case.

Possible Beta V0.1 failures include:

- a **false positive** that reports drift where no material drift occurred;
- a **false negative** that misses a meaningful substitution;
- an **ambiguous difference** for which more than one interpretation remains reasonable;
- a **qualifier distortion** introduced by DEDDA's own explanation;
- a **permission error** that overlooks or invents authorization;
- **intervention drift** in which DEDDA begins speaking for the user;
- **self-protective attribution** that moves the detector's failure onto its user;
- a **correction failure** in which DEDDA acknowledges a repair but repeats the operation;
- **valid-null avoidance** that manufactures findings to appear useful.

Each incident can leave a comparable footprint:

input

- expected operation
- observed DEDDA output
- user correction
- identified failure location
- changed process
- comparable retest
- persisted or returned?

A DEDDA error does not by itself invalidate DEDDA's purpose. How the app receives, records, tests, and later applies the correction reveals its process quality. Some failures may belong to the tool, some to its use, some to their interaction, and some may remain unresolved.

DEDDA must eventually inspect **DEDDA's own representation of correspondence**. The drift detector can drift. That is why corrigibility, temporal footprints, delta processing, valid nulls, and version history belong inside its development process.

1.11 Imagine the official version going viral :P

Imagine if the official version (stepping out of the beta phase) goes viral lol :P
The whole world can use it, hahahaha, instant comedy everywhere.

— John Kuhles, 17 October 2026

Millions of people could carry a small correspondence witness:

Person: "That is not what I said."

DEDDA:

"You supplied X.

The reply answered Y.

No connecting permission was found."

Replying A.I.:

"Thank you for that important—"

DEDDA:

"New sidetrack detected."

The comedy would come from making familiar substitutions plainly visible:

- "You removed *possibly*."
- "You answered the accusation rather than the question."
- "The user corrected this three turns ago."
- "That warning responds to a claim nobody made."
- "You negotiated for someone who explicitly disagreed."
- "No material drift detected—you two are arguing about the object itself." :D

At that scale, DEDDA could normalize one question:

Does the representation still correspond to the person or object it claims to represent?

Its public discipline would remain the same: show the correspondence path, preserve uncertainty, accept correction, return control, and inspect its own output. A DEDDA finding remains a corrigible finding rather than a final verdict.

Consented, privacy-protected patterns could also reveal recurring distortion mechanisms across many cases. One mismatch is a specimen. Many comparable specimens may expose a structural pattern. The pattern still needs provenance, qualifiers, and a correction route.

1.12 Commercial concept — “His Second Word”

John’s original commercial idea:

Mother & Father are proud of their baby son, who is starting to speak his first words, “mama,” in tears of joy... then a couple of days later, the father hears the baby saying “DEDDA,” and the father gets interrupted by an APP lol :P

— John Kuhles, 17 October 2026

The name itself becomes the plot mechanism because **DEDDA** sounds close enough to **Dadda** for a loving father’s interpretation to create the drift.

1.12.1 Scene 1

Mother and father lean over the crib.

Baby: “Mama.”

Mother bursts into tears of joy. Father hugs her.

Father: “His first word!”

1.12.2 Two days later

Father records the baby on his phone.

Baby: “DEDDA.”

Father freezes. His eyes fill with tears.

Father: “Dadda! You said Dadda!”

The phone gently chimes. A DEDDA card appears:

Possible correspondence drift

Source audio: “DEDDA.”

Later representation: “Dadda.”

No reliable permission from the speaker to resolve the ambiguity. :P

Father stares at the phone.

Father: “Are you seriously correcting me right now?”

DEDDA displays:

Motive unresolved. Phonetic enthusiasm detected.

Mother enters, watches the recording, and laughs.

Mother: “He was trying to activate the app.”

The baby points at the phone:

Baby: “DEDDA!”

1.12.3 Closing screen

DEDDA

Before somebody turns what was said into what they hoped was said.

Then, in smaller text:

Material drift—or a valid null. You remain the judge.

The commercial introduces DEDDA through affection, ambiguity, and a tiny act of representation drift. The father is emotionally motivated to hear “Dadda”; the joke needs no malicious actor.

1.13 Working principle

Keep the correspondence authentic enough that the participants can still find themselves in it—and keep the detector modest enough that it does not become another participant who replaces them.

1.14 Related phase-change concept

Rogue Ethics uses DEDDA’s qualifier, provenance, asymmetry, attribution, permission, delta, and correction-return operations to inspect a possible transition away from consensus ethics or ethics-washing. DEDDA does not presume the departure is benevolent or deep-ethical; it can report that the ethical direction remains unresolved or that no material movement was established.

Concept and subtitle: John Kuhles [parrot] [butterfly] [tree]

Public concept footprint: 17 October 2026

Primary illustrative source: [Graham Dot Zip on YouTube](#)