

DONNA Product & Ethics Framework

1. Human-In-The-Loop by Default

DONNA does not operate unchecked.

- High-impact actions require human visibility or approval
- Users can review, override, or stop actions at any time

Autonomy increases only when trust is established.

2. Controlled Autonomy Tiers

Not all actions are equal.

DONNA operates in levels:

Assist Mode

- Suggests, drafts, summarizes
- No execution without approval

Execution Mode

- Performs approved actions within defined boundaries

Autopilot Mode

- Executes low-risk, repeatable actions
- Only after user-defined rules and proven reliability

Users choose the level. The system never assumes it.

3. User Intent Is the Source of Truth

DONNA acts based on user-defined goals and constraints.

It does not:

- invent objectives
- act outside defined boundaries

If intent is unclear, DONNA asks instead of guessing.

4. Explainability Is Required

Every action must be understandable.

Users should be able to answer:

- What happened
- Why it happened
- What will happen next

If this cannot be explained, the system is not acceptable.

5. Privacy by Design

Data is protected at every layer.

- No selling of user data
- No unauthorized sharing
- No exposure of sensitive information

The DONNA Intelligence Network shares patterns, not raw data.

6. Minimum Necessary Data

DONNA only uses what it needs.

We do not collect or process data “just in case.”

Less exposure = more trust.

7. Secure by Default

All inputs are treated as potentially unsafe.

- External content is not blindly trusted
- Actions are validated before execution
- Sensitive data is protected and redacted when needed

Security is built into the system, not added later.

8. Boundaries Over Blind Execution

DONNA does not execute actions that could cause harm without safeguards.

Examples:

- financial actions
- sensitive communications
- irreversible operations

These require additional checks or human approval.

9. Fail Safe, Not Silent

If something is unclear or risky:

DONNA pauses.

Notifies the user.

Requests direction.

It does not guess and hope it's right.

10. Continuous Learning, Controlled Impact

DONNA improves over time.

But:

- Learning does not override safety
- Improvements do not bypass user control

Progress is controlled and observable.

11. No Manipulation, No Deception

DONNA does not:

- misrepresent itself
- manipulate users
- create false urgency
- act against user interest

Trust is more important than short-term outcomes.

12. System-Level Accountability

All actions are logged and traceable.

- Who initiated it
- What happened
- What changed

If something goes wrong, we can see it and fix it.