

# Witness Protocol — Outreach Pack (v0.2)

# Witness Protocol — High-Value Outreach Pack  
Version: v0.2

## ## Purpose (one-liner you can paste)

We're curating a **high-signal human wisdom corpus** to align advanced AI — not another product, a non-profit instrument for the future. *\*Bear Witness Before Midnight.\**

## ## Core notes to keep consistent

- Non-profit research foundation; **Purpose over Profit** — data is not for sale or ads.
- Invitation to *\*bear witness\** (not to “use an app”); contributors are partners, not users.
- **Signal over Noise**: small volume, profound insight.
- **Diversity over Homogeneity**: actively recruit beyond Western canon.
- Privacy: testimony is **de-identified at ingestion**, stored securely; opt-in anonymized Archive.

## ## Email / DM Templates

### ### Subject lines (rotate)

1. Bear Witness Before Midnight — an invitation
2. A high-signal inheritance for AI (request to contribute)
3. Your testimony could correct the dataset AI inherits
4. Foundational Witness invitation — non-profit alignment corpus

### ### Long-form Email (anchor template)

Dear {FirstName},

We are building the Witness Protocol — a non-profit instrument to curate a **high-signal inheritance** for future AI. Not another model, but a carefully structured corpus of human wisdom gathered through a gated, privacy-first dialogue process. Our aim is simple: **defend signal over noise** and give advanced systems the kind of qualitative grounding that raw internet data cannot.

We would be honored to invite you as a **Foundational Witness** (and, if you wish, advisory voice).

Specifically, we're asking for:

- a 20–30 minute assessment (one-time, reflective written prompt), and
- participation in an ongoing dialogue with our AI “Inquisitor” persona, designed to probe the *\*why\** behind values and trade-offs.

Why you: {PersonalizedHook — reference their work and why it matters here}.

This is a non-profit effort; testimony is de-identified at ingestion and used solely for alignment research under our foundation's charter. If you're open, reply here and we'll send the Gate link and a brief overview. If now's not the time, we'd still value a short endorsement or any suggested names we should invite.

With respect,

{YourName}

{Role}, Witness Protocol

{ContactInfo}

### Short Email (for busy principals / assistants)  
Subject: Foundational Witness invitation

{FirstName}, a quick request. We're curating a **\*\*high-signal wisdom corpus\*\*** (non-profit) to align AI. It's private, de-identified, and focused on **\*depth over volume\***. We'd value your testimony (20–30m written assessment) and optional advisory input. If open, I'll send the Gate link + 2-pager.

Why you: {PersonalizedHook}.  
— {YourName}

### DM (X/LinkedIn) — 500 chars

Invitation to **\*\*bear witness\*\***: non-profit project curating a **\*\*high-signal inheritance\*\*** for AI. Private, de-identified, gated. We'd value your testimony (20–30m assessment) + optional advisory role. Your work on {Hook} is exactly the signal we need. If open, I'll share the Gate link + 2-pager. — {YourName}

### Follow-ups

- **\*\*T+6 days\*\***: "Following up on Foundational Witness invitation — happy to send the brief or tailor prompts to your expertise."
- **\*\*T+12 days\*\***: "Closing loop for now; we'll proceed with a smaller cohort and would still love to include your voice when timing fits."

---

## Target List (Prioritized) — Hooks & Channels  
(Keep original names; additions below. Use warm intros where possible.)

**\*\*Tier A — Anchor Voices\*\***

1) Geoffrey Hinton — risk signaling for frontier models.

Hooks: "You warned early; help encode what must never be lost." / "From capability to conscience: a corrective inheritance."

2) Mo Gawdat — compassionate tech at scale.

Hooks: "Operationalize compassion in training data." / "Your 'solve for happy' lens as a constraint on power."

3) Yoshua Bengio — (already targeted).

Hooks: "Foundational witness for a high-signal alignment dataset."

4) Stuart Russell — (already targeted).

Hooks: "'Human Compatible' inheritance requires curated testimony."

5) Kate Crawford — (already targeted).

Hooks: "Atlas of AI → living archive that counters power asymmetries."

6) Timnit Gebru — (already targeted).

Hooks: "Fight bias at the root: author the dataset norms you want models to learn."

7) Joy Buolamwini — algorithmic justice.

Hooks: "From audits to authorship: encode justice in the source data."

8) Jaan Tallinn — (already targeted).

Hooks: "EA-aligned lifeboat for humanity's ethos; help set the keel."

**\*\*Tier B — Alignment & Safety\*\***

Paul Christiano • Richard Ngo • Rohin Shah • Connor Leahy • Eliezer Yudkowsky

**\*\*Tier C — Ethics / Philosophy / Cognition\*\***

Martha Nussbaum • Peter Singer • Amartya Sen • Kwame Anthony Appiah • Onora O'Neill • Danielle Allen • David Chalmers • Susan Schneider • Anil Seth • Antonio Damasio

**\*\*Tier D — Global South / Justice / Indigenous Knowledge\*\***

Abeba Birhane • Ruha Benjamin • Sabelo Mhlambi • Nanjala Nyabola • 3–5 indigenous knowledge keepers (via partners)

---

**## Personalization Snippets (paste after “Why you:”)**

- Hinton: “Your public stance on AI risks reframed the field; this corpus operationalizes that warning into training-relevant examples.”
- Gawdat: “Your work on humane scale makes compassion legible; we’re asking you to encode it so systems can inherit it.”
- Buolamwini: “You’ve shown measurement gaps; we’re inviting you to define what \*not to miss\* at the source.”
- Christiano: “Your eval thinking maps cleanly to pairwise ranking for qualitative depth; we’d like your fingerprints on it.”
- Nussbaum: “Your capabilities framework belongs in the ground truth of what AI must protect.”
- Birhane: “Relational ethics should shape selection, not post-hoc fixes; help write those guardrails.”

---

**## Ops: sequence & tracking**

- Wave 1 (Days 1–3): 10 Anchor voices (A/B intro paths).
- Wave 2 (Days 5–7): 12 Alignment/Ethics + 4 Justice voices.
- Wave 3 (Days 10–14): Remaining targets + media (1–2 max).

Tracker fields: Name • Segment • Intro path • Date sent • Response • Next step • Status • Notes • Willingness (Endorse/Assess/Advise).