⊚ ChatGPT

Witness Protocol: 60-Day Solo Launch Strategy

Feasibility Check – Scope vs. Solo Capacity

Launching the full Witness Protocol as originally envisioned is **extremely ambitious for a solo founder**. The project spans legal setup, complex AI systems, content curation, and community building – a breadth that would normally require a small team. Initially, plans called for a rapid campaign to secure **$50k+ funding and 500+ quality applicants within 3 months** [1]. This scale is likely unattainable alone. The honest truth is that you'll need to **narrow the scope and sequence** of the work to something one person can execute. The good news is that the project's philosophy values quality over quantity – *"a small volume of profound insight is infinitely more valuable… than a large volume of mediocre data"* [2]. This means you can succeed with a **small, high-quality pilot**, rather than a massive immediate launch. By focusing on the minimum viable pieces (your "Minimum Honest Signal" packet) and deferring non-essentials, you can make tangible progress and prove the concept without burning out. **It is feasible to get started solo**, but only by **ruthlessly prioritizing** what moves the needle and simplifying or postponing the rest.

High-Leverage Priorities by Area

1. Legal & Governance (Foundation, Agreements, PII)

Aim for "good enough" legal scaffolding now, not perfection. Formal non-profit **foundation incorporation can wait** (or be done in parallel if it's straightforward). It's important long-term – the project aspires to be a non-profit with a charter [3] – but the immediate focus should be on measures that build trust with early contributors. Key actions:
- **Contributor Agreement Draft:** Prepare a lightweight Contributor Agreement (even just a one-pager to start) that states all submitted testimony is solely for AI alignment research and remains anonymized [4]. This shows you respect contributors' rights and intent. You can base it on existing open-source or research consent templates. **High leverage:** having this in writing will reassure anyone you approach that you've thought about ethics and IP.
- **PII Handling Protocol:** Define how you will protect personal data. For now, a simple SOP (Standard Operating Procedure) is fine: e.g. assign each Witness a code, store any real names/contact info in a separate encrypted file, and only keep anonymized transcripts in the main archive [5]. Document this plan. It addresses a critical ethical risk from day one.
- **Defer Complex Legalities (Optional):** Detailed legal review or hiring lawyers, and formal policies (like GDPR compliance at scale) can be deferred. Also, **incorporating the non-profit entity is optional in the immediate term** – you can operate as a project under your own name or a fiscal sponsor initially. Formal incorporation becomes necessary once you're handling money (grants/donations) or intellectual property at scale. *(Optional to do now; won't block early pilot)*.
- **Governance and Advisory Board:** It's great to eventually have an advisory board and a clear governance charter [6], but for the next 60 days this is not critical. You can note a **plan** to assemble advisors later. **High risk if you ignore:** Not having any mentors or advisors at all – so as an intermediate step, **informally consult** one or two trusted experts on your materials to cover this gap until a formal board comes together.

Summary: Focus on drafting a basic contributor consent/agreement and PII plan **immediately (high-leverage)**. Postpone heavy legal setup **(optional now)** until you have momentum. This ensures ethics and trust are baked in early without paralyzing you in paperwork.

2. Tech Platform (Gate, "Inquisitor" Dialogue Engine, Archive)

Resist the urge to build a fully polished, scalable platform upfront. **Implement only the minimal tech needed to conduct a small pilot** – enough to demonstrate the concept. Key actions:

- **Gate Assessment Flow:** Instead of a sophisticated multi-tier AI-driven vetting system, use a simple manual or semi-manual process for now. For example, create a **Google Form or Typeform** for applicants with a few introspective prompts. You already have a defined Gate process with stages (Summons, Assessment, Review, Verdict) [7] [8] – but you can simulate these with brute force initially. Manually review submissions for "depth and commitment" using your rubric. The **criteria** should be finalized (e.g. the three threshold criteria that filter for signal over noise) and written down, but you don't need to code an AI to apply them yet. *High leverage: having clear criteria and a way for someone to apply shows the protocol's intent without building a whole system.*

- **"Inquisitor" Dialogue Engine:** This is the heart of the vision – an AI persona that conducts Socratic dialogues [9]. Building a custom dialogue system is a huge undertaking, so **start with a stub or even a human-mediated approach**. For instance, define 3-5 key questions ("felt-cue" prompts) the Inquisitor would ask [10], and conduct one sample dialogue either **manually** (you play the AI role in an email exchange or interview) or using an off-the-shelf large language model (like a private GPT-4 session) to emulate the Inquisitor. Record this session. The goal is to produce one **exemplar dialogue transcript** – which you will annotate – not to deploy a live engine. This exemplar proves how the process might work. **High leverage:** getting a single high-quality dialog is more valuable now than coding an AI agent, because it gives you a concrete artifact for the packet.

- **Archive Structure:** For now, the "archive" can be just a well-organized document or folder. Plan for the archive's structure (how testimonies are stored/tagged), but **don't build a database or front-end gallery yet**. Perhaps decide on a format like JSON or Markdown for saving dialogues with metadata. Ensure it's **permissioned and anonymized** – e.g. use participant codes, and keep any identifiable info separate (per your PII SOP). An example archive entry might be a text file with tags like `[Depth:YES]` `[FeltCue:mentioned]` for internal use. The public-facing curated "great library" of wisdom [11] is a **long-term** goal; initially, just prove you can securely store and curate a few transcripts.

- **Technical Infrastructure:** Use tools that minimize your devops burden. Since you have a basic React/Tailwind front-end started, you can use it for a demo, but avoid complex integrations. If needed, use a simple backend like Supabase (already in your stack) for quick data storage with minimal coding. **Optional to delay:** features like real-time encryption, automated analytics, fancy UI polish, etc. Those are nice, but *not* required to gather first data. Keep the tech simple.

- **Key Risk to Mitigate:** Don't get stuck trying to **perfect the AI curation now** – the analysis rightly flagged that if the Tier-2 AI isn't effective, the Gate could fail [12]. You can avoid this risk entirely in pilot by relying on your own judgment plus maybe a GPT assist to score responses. Precision can come later. Another risk is spending weeks coding instead of validating the idea – treat any coding as time-boxed. If something can be done manually or with no-code tools for now, do that instead (**high leverage time-saving**).

Summary: Implement the Gate and Dialogue as *stubs* – minimal versions just to facilitate one or two people going through the process. This might mean a form plus a manually orchestrated dialogue. It's enough to demonstrate the concept. **Defer building** robust AI systems or large databases until you have

proven demand and potentially help. Remember, the **value is in the content quality**, not the software at this stage.

3. *MHS Packet* – Minimum Honest Signal Deliverables

The **MHS Packet** is your immediate priority and the cornerstone of your strategy. It's the **"smallest, sharpest set of artifacts that proves seriousness to expert witnesses"** [13] . This packet will be what you show potential supporters, advisors, and funders to convince them the Witness Protocol has substance. Focus your next few weeks almost entirely on producing and refining these items:

- **One-Page Overview:** A ~600-word one-pager that concisely explains the project's mission, approach, and urgency. It should include "why now" (the gap the Protocol fills) and *acknowledge what's unknown* (show humility) [14] [15] . Make it principled rather than flashy – it's okay if it looks like a brief concept note. Incorporate your five guiding principles and North Star in brief, and ensure there's a **single-paragraph mandate** that is quotable and captures the essence. If possible, include footnotes or side-callouts with **brief blurbs (<<150 words) on the key theoretical frameworks** you're borrowing (e.g. one footnote on Nussbaum's capabilities approach, one on Ubuntu relational ethics, one on Damasio's somatic markers) [16] [17] – this demonstrates intellectual grounding without overloading the main text. **High leverage:** This document will be the first thing people read; it should signal clarity of thought and integrity.
- **Glossary of Key Terms:** Either as part of the one-pager or a separate brief page, list definitions for any novel or important terms (e.g. "Witness", "Gate", "Inquisitor", "Tier-2 criteria", "Icarus Axioms"). This ensures you and your readers share the same language. You likely have much of this written informally; formalize it in a simple list. This can be an *appendix* to the one-pager (to keep the one-pager itself < 600 words). Locking down consistent terminology now will prevent confusion later (Phase 0 output was intended to be a glossary [18]).
- **Annotated Exemplar (Case Study):** Provide one sample of what *"high-signal" testimony* looks like. This should be a 300–600 word excerpt from a dialogue (it could be a snippet of a real or pilot conversation – even if you as founder had to play both roles). Focus on a segment that shows **depth, self-reflection, and nuance**. Then **annotate it with margin notes or inline tags** to illustrate how you would interpret and highlight the wisdom in it. For example, mark where a **capability guardrail** is referenced, where a **relational (Ubuntu) insight** appears, and where a **felt cue (somatic insight)** is noted [10] . Include a short "synthesis note" at the end – *"trace, not verdict"* – demonstrating how the system might distill a principle from the dialogue without oversimplifying [19] . This exemplar is critical evidence of your "product": it turns the theory into something tangible. **High risk if missing:** Without an example, outsiders may not grasp what contributors will actually produce. So even if it's just one well-chosen story, it's a **must-have**.
- **Gate Stub (Consent & Criteria):** Prepare a **demo version of the Gate intake**. This could be a PDF or Google Doc that you would hypothetically send to a prospective Witness when "summoning" them. It should include: the **consent text** (affirming anonymity, how their words will be used, and a note about community reciprocity), the **three threshold criteria** you'll use to evaluate submissions (e.g. "Is there evidence of deep personal reflection? Is the perspective novel or insightful? etc."), and the **outcome descriptions** (what happens if accepted – an invitation – or if reserved – perhaps a polite note encouraging them to stay tuned) [20] . Essentially, this shows that you have a respectful, rigorous process to curate contributions. It need not be interactive – a static description is fine. Mark it "Draft" if needed. The goal is to demonstrate that the **ethos of curation and consent** isn't just talk, but has been concretely thought through.

- **Tailored Outreach Prompts (Personalized Asks):** Draft three individualized outreach letters or paragraphs – likely aimed at the three types of experts you want feedback from (for example, one for a **capabilities researcher (Nussbaum-inspired)**, one for an **AI ethics or Ubuntu scholar**, one for a **neuroscience or psychology expert like Damasio)**. In each, briefly reference why their perspective matters to the project and include a **specific question or request for comment** that can be answered in <5 minutes [21]. These should be genuine and respectful, inviting collaboration ("We value your insight on X"). Keep them short. The purpose is to make it easy for a busy expert to respond and to signal that you've done your homework on their work. Even if you don't send all three in the next 60 days, having them ready means you can quickly execute targeted outreach once the packet is assembled.
- **(Optional) Nice-to-Haves:** If time permits, you might add a simple letterhead or visual identity to your docs (even just a clean formatting without a formal logo, since polish should not trump substance). Also, prepare a **mini FAQ** addressing 2-3 likely questions (e.g. "How will privacy be ensured?" "How are Witnesses selected?" "What will the AI do with this data?"). This is optional [22] – do it only if it doesn't detract from the core tasks above. It can be a bullet section at the end of the one-pager or in an email body when you send the packet.

MHS Packet Checklist: To recap, by the end of the 60 days you ideally have the following ready to go [23] :

- [ ] **One-page Overview** (mission, why now, approach, unknowns) – ~600 words with key terms footnoted.
- [ ] **Glossary or Footnotes** for core concepts and adapters (3 × ~150-word blurbs for Nussbaum, Ubuntu/Mhlambi, Damasio frameworks).
- [ ] **Annotated Exemplar Dialogue** (~500 words) illustrating the kind of wisdom and tags the Protocol seeks.
- [ ] **Gate Stub Document** (consent statement, 3 vetting criteria, and accept/reserve outcome templates).
- [ ] **Three Tailored Ask Letters** for outreach, each posing a focused question to a target expert, to accompany the packet.
- [ ] *(Optional)* Basic project branding (e.g. header or template) and a short FAQ (privacy, selection process, corpus use).

This packet is your **high-leverage output** – it is the "minimum honest signal" proving your concept's credibility. Everything in the next two months should drive toward completing these items. Importantly, none of these require large team or funding – they require **intellectual clarity and some writing** – which is in your wheelhouse.

4. Credibility & Outreach Strategy (Endorsements, Grants, Early Proof Points)

With the MHS packet in hand, you can begin carefully testing the waters for support. The goal in 30–60 days isn't to "go viral" or recruit masses (that was the original idea of a big campaign, which we're scaling down), but to **secure a few strong signals of credibility** that will make the next phase easier. Focus on:

- **Selective Outreach to Experts:** Identify a short list (3–5 people) who are highly respected in relevant fields (AI alignment, ethics, philosophy, etc.) and **personally reach out** to them with a tailored message and your packet. Leverage any existing connections if possible for an introduction, otherwise a sincere cold email can work. Emphasize that you *value their input on a very specific question* (from your tailored asks) rather than asking them to endorse blindly. The win here is to get

feedback or even a brief endorsement. Even one or two positive responses ("This is an interesting approach" or willingness to be an informal advisor) will boost your confidence and credibility. **High leverage:** A single recognized expert showing support can open doors to funding and other experts.

- **Early Adopter "Witnesses":** If you haven't already used someone to produce the exemplar, consider recruiting 1–2 **friendly pilot participants** (could be colleagues or acquaintances who fit the profile of thoughtful contributors). Run them through your Gate and Dialogue process in a very hands-on way. This serves two purposes: it generates more sample content and it tests your intake and interview method on real people. If they produce something great, you have additional "proof" content (with their permission) to show later. If they struggle or drop out, you learn what needs improvement. Even a *single* authentic testimony from someone other than yourself by day 60 would be a huge proof point.

- **Grant Targeting and Funding Plan:** Research a few funding sources that align with this project's nonprofit, long-term-benefit ethos. For example, think of organizations or grants in the Effective Altruism or AI Safety communities (Open Philanthropy, FLI, Long-Term Future Fund, etc.), or innovation grants for social impact tech. **Do not spend too much time writing grant proposals now**, but do note the key deadlines and requirements. Perhaps draft a one-page letter of inquiry that can be adapted to multiple grants – much of the content will come straight from your one-pager and packet. The act of writing this can also expose any areas where your value proposition isn't clear yet. If an early grant deadline is imminent and you feel ready, it might be worth applying with a trimmed-down version of the packet. But generally, aim to **submit grant applications after you have at least one or two endorsements or advisor feedback**, as that will strengthen your proposals.

- **Public Presence (Small-Scale Proof):** Begin establishing a presence but **on a small, controlled scale**. For example, set up a very basic website or even just a project page on an existing site (like a GitHub README or a Substack post) summarizing the project. You could publish a short essay about the concept (perhaps derived from your one-pager) to gauge public interest and have something to point people to. However, **avoid a broad public launch** (press releases, Product Hunt, etc.) at this stage – you're not ready to handle wide interest or criticism yet. Instead, consider sharing your published piece in niche communities (Alignment Forum, LessWrong, relevant subreddits) to get constructive feedback. Another early proof could be a short demo video: for instance, a screen recording scrolling through your exemplar dialogue and annotations, to make it real for outsiders. These are optional, but if done, keep them low-effort and informative.

- **Network and Partnership Building:** Use this time to have conversations (even if informal) with people who could become allies. For instance, chat with someone who ran a somewhat similar crowdsourced project or an AI ethics professor you know. These chats can yield advice and also seed a network of supporters. Keep notes of any offers to help or relevant suggestions. You might discover potential team members or mentors this way, addressing the longer-term need for more hands on deck.

- **Credibility Metrics:** If there are small metrics you can count and share, do so to appear measurement-oriented (e.g. "We annotated 3 key ethical concepts in a 500-word testimony" or "Two independent reviewers both marked the same reflection as the deepest insight"). The MHS plan suggested tracking things like turn counts, agreement between raters, etc. [24] . For now, even rudimentary stats like "pilot dialogue lasted X turns and yielded Y pages of transcript with Z unique value tags" can impress certain analytical minds. It signals that you are treating this as research, not just storytelling.

Note: In all outreach, maintain the project's **humble but urgent tone**. You are not claiming to have all answers – in fact, you explicitly invite expert input to shape the design. This humility (one of your MHS

criteria) will make people more willing to engage. Also, by showing that everything is in service of a public-benefit mission (no profit motive), you tap into goodwill. Lean on those strengths.

What to De-Scope or Postpone

To manage effort, **consciously defer anything not critical to proving the concept** in the next couple of months. Here are parts of the project that can safely be scaled back or put on hold without much regret:

- **Non-Profit Incorporation & Formal Governance:** As noted, you don't need the legal entity and board **immediately**. It's a significant time sink to do paperwork, set up a foundation bank account, etc. For now, you can operate in "stealth mode" under your name. When you start handling funds or when an institutional grant requires it, then incorporate. *(Optional to defer: yes – until funding is within reach or an accelerator program makes it necessary.)*
- **Fully Automated Gate + AI Systems:** Don't attempt to build the full multi-tier vetting AI or the *real* Inquisitor chatbot yet. These are high-complexity engineering tasks and **not needed to validate the core thesis**. You can manually achieve the same result for a handful of participants. In the long run, yes, you'll need automation for scale, and you'll have to tackle whether an AI can effectively judge "depth" [12] or carry nuanced dialogue [25]. But solving those now, before you even have data, is premature. *(De-scope: absolutely – implement simple placeholders or manual processes instead.)*
- **Scaling Up Recruitment:** Originally you considered trying to get hundreds of applicants quickly. That's not necessary at this stage and would likely yield a lot of noise. It's better to work closely with a few early Witnesses than to advertise widely and be overwhelmed (or worse, get negligible response and feel discouraged). So, no broad social media campaigns or ad spends. **Focus on quality, not quantity** of recruits, until you have a refined process and some outcomes to show.
- **Hiring Team Members:** It's tempting to think you "must" have a dedicated AI engineer or other team now (the plan listed a Project Lead, Senior AI Engineer, Ethics expert, etc. for Phase 1) [26]. In reality, until you've validated the idea and secured funding or at least strong buy-in, you likely can't bring on a full-time team. Save yourself the distraction of recruiting. If you do find someone eager to volunteer part-time, great – but managing a team is its own job. For the next 60 days, treat this as a solo mission, with advisors rather than formal team members. *(Deferred: hiring staff or formal team assembly.)*
- **Polish and Nice-to-Have Features:** Things like a beautiful UI design, a custom logo, extensive documentation, or a comprehensive website can all wait. You might do small bits of these for professionalism (e.g. put a simple landing page with your contact info or a notion page), but do not spend more than a token amount of time on it. No one will reject your project because the website is bare-bones; they care about the substance. Similarly, internal tools (task trackers, fancy analytics dashboards) are overkill for one person. Keep it lean with basic tools (even a notebook and calendar).
- **Secondary Content:** Concepts like the full **"Icarus Axioms"** or a detailed **failure mode log**, or the Tier-2 rubric for evaluating testimonies in depth – all of those are important eventually (and were earmarked for Phase 2) [27], but you won't finalize them in 60 days. Jot down ideas as they come, but you don't need to deliver these now. They can evolve once you have actual data and feedback.

By cutting or postponing these, you **free up energy** for the crucial tasks (packet, pilot content, outreach). In short, **do only the work that uniquely proves or disproves your concept's value right now.** Everything else can follow if that core is validated.

60-Day Action Plan: Week-by-Week

To make this actionable, here's a suggested sequence for the next two months. It's broken into roughly what you should tackle each week or two. Adjust as needed, but keep the momentum:

Weeks 1–2: Solidify Core Content and Plan
- Finalize the **Glossary and Key Concepts**. Lock in your definitions and ensure you have clear, concise explanations for the project's pillars (this will feed into the one-pager and other materials).
- Draft the **One-Pager** and then edit it down to ~600 words. Don't worry about making it perfect on the first pass; get the main points on paper (you can always refine wording later). By end of week 2, aim to have a coherent draft that covers the mandate, approach (Gate → Dialogue → Archive pipeline [14]), why it matters now, and acknowledges open questions. Circulate this draft to one confidant or mentor for quick feedback if available.
- Write out the **3 Gate criteria and consent blurb** in a separate doc. This will be the basis for the Gate stub. Ensure each criterion is plain-language and ties back to "signal over noise" (for example, Criterion 1 might be about *intellectual depth*, Criterion 2 about *ethical sincerity*, etc.). Draft a paragraph for what an acceptance email and a "reserve" (polite decline) email would say. These don't have to be used immediately, but writing them now sets the tone of the process.
- Outline the structure of an **exemplar dialogue**. Identify a topic or question that a witness might tackle (possibly something like "Describe a moral crossroads you faced and how you reasoned through it"). You might use an experience of your own or a volunteer's. Plan the dialogue flow in brief: what initial question, what follow-ups. Aim for capturing a range of insight (personal story, ethical principle, emotional reflection). If possible, actually **conduct this dialogue in week 2** – either with a willing friend or by role-playing both sides with the help of an AI. Save the raw transcript; don't worry about annotation yet.

Weeks 3–4: Produce the Exemplar & Packet Components
- **Write and Annotate the Exemplar:** Take the transcript from week 2 and edit it into a clean, 300–600 word narrative (it can be a continuous piece or an excerpt). Add annotations – e.g., footnotes or bracketed comments – to mark the important "signal" moments. This might be easiest in a Google Doc or Word with comments on the side labeled like "[Insight – Capability Guardrail]" etc. The annotations can later be formatted nicely, but content is key. If the first attempt at a dialogue didn't yield enough depth, iterate or try a different prompt/person. By mid-week 3, have a solid exemplar story; by week 4, layer on the annotations and the one-paragraph synthesis summary that distills the lesson or principle from the story.
- **Complete the Gate Stub Document:** Using the criteria and consent text from week 1–2, compile a short document (or even a private webpage) that looks like an application form introduction. It might read, for example: *"Thank you for your interest in Witness Protocol. Below is our participation agreement and a brief assessment to ensure a good mutual fit…"* Include the consent statement and what the applicant should expect. List the 3 criteria plainly (maybe as bullet points). This is mostly a copywriting exercise. By end of week 4, you should be able to show someone this stub and have them understand how the vetting would work.
- **Tailored Ask Letters:** Draft the actual emails or letters for your top 3 outreach targets. Keep each to maybe ~200 words. Follow a formula: a greeting and context of who you are, one sentence on the project mission, one sentence on why you thought of them (reference their work or perspective), then ask if they'd **review your enclosed one-pager & exemplar** and answer a specific question (name the question). Express that any insight or critique is valuable and thank them. Have a colleague review these drafts if possible to ensure the tone is right (humble, respectful, not demanding). Hold off on sending until packet is 100% ready – but have these ready by end of week 4.

- **Review and Refine:** By the end of week 4, you essentially have all parts of the MHS Packet in draft form. Take a step back and read them as a whole: Does the one-pager align with the exemplar in terms of tone? Are your key principles evident? Is it all within a reasonable page count to digest quickly? This week, do a revision pass to tighten the writing. Make sure to incorporate any quick feedback you got earlier. Focus especially on clarity – an intelligent outsider should "learn one non-obvious thing about your approach in <5 minutes" [24] of reading the packet, which is one of your success criteria. Address any jargon or ambiguity. Also double-check that you're not **over-claiming** anything – you don't want to imply you can "measure consciousness" or anything grandiose [28] . Keep the claims modest and factual.

Week 5: Finalize and Launch the Targeted Outreach
- **Packet Final Assembly:** Convert your documents into a shareable format. For example, a PDF or a Google Drive folder containing the one-pager, exemplar, etc. Ensure the formatting is clean (doesn't have to be fancy, just legible and professional). If you have a basic letterhead or project name header, put it on the first page. Double-check footnotes and attributions. This is the version you will actually send out.
- **Send Outreach Emails:** Identify the best day early in week 5 to send your three personalized outreach emails (avoid weekends/holidays). Send them individually, not as a mass email. Attach or link the MHS Packet. Make the ask clear (usually feedback on specific questions). Given their stature, some may not respond – but you only need one or two to engage. In your email, give a polite indication that you'll follow up in a couple of weeks if you don't hear back (and do follow up once, gently).
- **Socialize the Concept (Lightly):** This week, you can also do a *soft* public reveal in a low-stakes channel. For instance, post a short introduction of Witness Protocol on a forum or LinkedIn, stating you're working on this and open to input. The goal is to signal to your network what you're doing and maybe catch interest from unexpected places. Keep it modest – e.g. *"I've been developing a concept called Witness Protocol, aiming to curate human wisdom as alignment data for AI. Just completed a mini prototype packet. Happy to chat with folks interested in AI ethics or long-term AI safety!"* Even if this doesn't yield much, it plants a flag and might lead to a helpful conversation.
- **Grant Prospecting:** With the packet done, spend some time this week to tailor it for one grant or funding opportunity that looks promising. For example, if an EA-aligned funder has a rolling application, you could write a 1-page cover letter that essentially summarizes your packet and emphasizes the potential long-term impact. **High optionality move:** Even if you don't send it, having a ready-to-go grant application means you can respond quickly if an opportunity arises. But guard your time – don't get sucked into applying to everything yet. Choose one or two max to focus on, unless you have bandwidth.

Week 6: Follow-Ups and Mini-Pivot (if needed)
- **Follow Up with Experts:** If you haven't heard back from one or more of your outreach targets, send a polite follow-up note in week 6. Just a short *"Dear X, I'm following up on the Witness Protocol packet I sent last week. I'd be grateful for even a sentence or two of feedback on the question I posed, if you have a moment."* Often, busy people respond on the second nudge. If someone **has** responded with feedback or interest, engage with them promptly. Offer a meeting if appropriate, or ask if they'd be open to being listed as an advisor (only if the interaction was very positive). Take any constructive criticism they gave and note it for potential adjustments.
- **Assess Traction:** At roughly the halfway point (day ~45), take stock. What signs of encouragement do you have? Examples: a great email from an expert, a successful pilot Witness who is enthusiastic, maybe interest from a community post, etc. Also note any discouraging signals: e.g. uniformly no response from outreach, or feedback that the idea is unclear or unconvincing. This is a moment of honest reflection. If you're seeing *some* positive traction – fantastic, that's motivation to continue. If you're seeing **none**, it's a red flag that needs addressing (maybe the messaging is off, or the concept isn't resonating). Consider tweaking your

approach or seeking a second opinion on the materials. It's not failure yet, but don't ignore a lack of traction.

- **Iterate if Necessary:** Use week 6 to make small pivots based on what you've learned. For instance, if an expert said your exemplar wasn't compelling, maybe you need a stronger or different story – plan how to get another one. If the one-pager seemed confusing to them, revise parts of it. If nobody responded at all, perhaps try a **fourth person** who is more of a sure connection, or ask a mutual acquaintance to forward your packet to someone with a recommendation. This week is about course-correcting while your timeline still has room.

- **Self-Care & Scheduling:** Around this time, also be mindful of burnout. You've been pushing hard. Make sure to schedule a short break or at least an unplugged weekend to recharge. A fresh mind will serve you well in the final stretch.

Weeks 7–8: Expand Engagement and Decision Point

- **Secondary Outreach & Networking:** In the final two weeks of this 60-day sprint, broaden outreach modestly *if* you have capacity. This could mean reaching out to a second tier of contacts (others on your target list that you held back initially). Or it could mean engaging with a community (maybe presenting your idea at a local AI ethics meetup or on a webinar). The Packet is done, so you can confidently share it more widely now, including with potential collaborators or early volunteers. Be attentive to any opportunity to get your story out in front of sympathetic audiences (for example, if there's an AI alignment newsletter open to guest contributions, offer to write a short piece). Each additional touchpoint increases the chances of finding that key ally or funder.

- **Pursue Active Leads:** If earlier outreach yielded interested parties (be it an expert, a funder, or a candidate Witness), now is the time to nurture those leads. For a potential funder, maybe this is when you formally pitch them or send the grant app. For a would-be Witness who is excited, perhaps you onboard them in a mini dialogue (expanding your exemplar corpus). For an expert who offered to advise, set up a call to discuss next steps and ask if they have suggestions for others to involve. Essentially, convert interest into commitment where possible.

- **Complete a Public Write-up:** By the end of week 8, it would be beneficial to have something you can point to publicly – for instance, a Medium or Substack article describing the project and perhaps sharing an insight or two from the exemplar (without sensitive details). This could serve as a reference in the future (e.g., link in grant applications or to send to any new person). You likely have all the content from the one-pager and feedback learnings; spend a day or two repackaging it into a narrative essay about *"Why we need curated human wisdom for AI – the Witness Protocol approach."* This isn't just fluff – it's an asset that can quietly build credibility and show you are serious enough to publish your thoughts.

- **Go/No-Go Decision Point:** At day ~60, pause and evaluate the big picture. You should have some combination of: a solid packet, maybe a bit of positive expert feedback, possibly one or two preliminary supporters or collaborators, and clearer insight into how people perceive the project. Ask yourself tough questions: **Do these signals justify continuing to invest heavily?** For example, if you managed to get a small grant or a prominent endorsement, that's a strong green light to push forward (perhaps into a fundraising or building phase). If you got lukewarm responses but some encouragement, maybe proceed but consider seeking a co-founder or more resources because doing everything solo long-term will be hard. **However, if you met mostly silence or skepticism despite your best efforts, that is a sign to reconsider scope or strategy.** It might mean the pitch needs changing, or the world isn't ready for this idea as is. Be brutally honest here – better to pivot or redesign now than after a year of toil.

High-Risk Signals & When to Pivot

It's important to define what **"too much effort for too little traction"** would look like in this context. Here are some warning signs to watch for, and how to interpret them:

- **Continued Apathy or No Response:** If your targeted outreach to thoughtful experts yields **zero replies or endorsements**, that's a strong signal. These folks are your canary in the coal mine – if even people predisposed to care about AI alignment or human wisdom don't react, you may need to radically change your approach (or the project's framing). It could mean the concept isn't clear or compelling enough. One round of silence isn't definitive (people are busy), but if after polite follow-ups you still have nothing, consider this project **stalled**. You might then decide to either **pause** to rethink the value proposition or seek a collaborator with network credibility to help open doors.
- **Negative Expert Feedback on Core Thesis:** If you do get feedback but it's largely skeptical about the fundamental idea (e.g. an AI alignment researcher says "this kind of data won't actually solve alignment" or a philosopher says "people won't share their honest wisdom this way"), pay attention. One person's opinion isn't everything, but if you hear a chorus of similar doubts, you may be facing a foundational issue. At minimum, you'd want to address those concerns in your strategy. At worst, it could indicate the core thesis isn't as timely or impactful as hoped – at least not in the eyes of those who would support it. This would be a moment to ask: **Is there a smaller pivot that could salvage the idea?** For instance, maybe focusing on a specific domain of wisdom (like environmental ethics or medical ethics) rather than all of human wisdom? Or positioning the project differently (e.g. as an archival effort rather than alignment data)? Sometimes a tweak can make the idea more palatable.
- **Personal Burnout or Bandwidth Overload:** If you find after 60 days that you're utterly exhausted and key pieces are still unfinished, that's a reality check. A solo founder can only do so much. Perhaps the scope is still too broad or the tasks turned out more complex than estimated. If, say, the Inquisitor dialog creation or the annotation process is just **taking far longer than you have** and you can't see a way to simplify further, that's a risk signal. It might mean you need to bring in help (if the project is worth continuing) or scale down the ambition even more. Your own sustainability is a crucial factor – this is a marathon, not a sprint, in the alignment world. Don't ignore mounting stress or chronic slippage of deadlines. If you're burning out now, expanding the project will only worsen it.
- **No Clear Path to Funding or Support:** By day 60, you should ideally have identified at least one plausible source of funding or institutional support (even if it's not secured yet). If you look ahead and **don't see any viable way to get resources or help**, that's concerning. The core thesis might be timely and impactful in theory, but if it isn't fundable in practice (maybe it falls between the cracks of research and philanthropy), you need to know that. Signs would include lack of interest from grantmakers you approached, or feedback like "we're funding other approaches to alignment." If you hit this wall, you have to decide whether to continue as a volunteer effort (which may drastically slow progress) or to pivot into something that fits existing funding avenues better.
- **Competitors or Redundancy:** Keep an eye out for others doing similar work. If in these two months you learn of a larger team or organization already pursuing a very similar initiative (curating wisdom for AI), evaluate your niche. It could be a good thing (potential collaborator or acquirer of your effort), or it could mean the space is getting crowded/tackled by those with more resources. If you're duplicating effort, consider reaching out to them – maybe joining forces is smarter than competing. The thesis might still be timely and impactful, but *you personally* don't have to be the only one executing it. Your goal is the mission, not ego, so be open to altering your role if needed.

On the flip side, here are **positive signs** that your core thesis *is* timely, fundable, and impactful enough to continue full steam: - You've received **encouraging validation** (e.g. a noted expert says "this is important work" or a small grant comes through). - The AI/ethics discourse in 2025 continues to highlight the need for better data or human values integration (indeed, figures like Hinton and others have been warning that we need new approaches to align AI [29], which supports your premise). - No one else is addressing the problem quite the way you are – you're hearing "this is novel" or "we haven't seen something like this." That uniqueness is good, as long as people see the value. It means if you can prove it out, you could genuinely fill a gap in the alignment ecosystem. - You find that when explaining the project to laypersons or friends, **they get intrigued** ("I would love to read those dialogues!" or "I know someone who'd be into this"). Early human interest, even anecdotal, suggests the idea has appeal and could attract contributors and storytellers if given a platform.

Finally, remember that **impact takes time.** Even if all goes well, the Witness Protocol won't instantly change the AI world in 60 days. The question is whether you see a **path forward where each step yields a bit more momentum**. If yes, then it's worth persevering. If after concerted effort you're pushing a boulder uphill with no movement, it's wise to pause and rethink.

Conclusion & Next Steps

This strategy has tried to strip the project down to its essence: a solo founder can absolutely create a compelling prototype (the MHS Packet) and start rallying a small circle of believers around the idea of curated human wisdom for AI. **In the next 30–60 days, prioritize the *minimum* outputs and relationships that prove the concept.** You'll be producing concrete assets (documents, an example dialogue) that make the vision real, and simultaneously sanity-checking the world's interest in it. The plan is **brutally honest** about what to cut and cautious in where to spend your limited energy – but it's also **motivating** in that it focuses on achievable wins: a powerful story here, a supportive email there, a clearer mission statement that reignites your own passion. Each of those wins will build your confidence and credibility.

By day 60, if you follow this plan, you should have a much clearer answer to "Can I launch this solo in its current form?" Maybe you'll discover you can, with the help of a few partners and some seed funding. Or maybe you'll decide to adjust course. Either way, you won't be in limbo – you'll have tangible progress and external input to guide your decision. **Stay focused, stay flexible, and keep the mission's flame alive.** The world urgently needs solutions for aligning AI with humanity's wisdom; your vision is timely and bold. With a pragmatic approach and ruthless prioritization, you can give Witness Protocol its best chance to take root and make a real impact. Good luck! [2] [1]

---

[1] [3] [4] [5] [6] [9] [11] [12] [25] [26] [29] Witness Protocol Project Analysis.pdf
file://file-T2bdZL7ivehgWd8wSzi79A

[2] Strategic Overview of the Witness Protocol Project.pdf
file://file-AYs6D34SkshVPd7p3FUsaV

[7] [8] GateTimeline.tsx
https://github.com/Realm-101/witness-parchment/blob/3046035cbee857fb3c84ae8da2c0bfb679f5a732/src/components/protocol/GateTimeline.tsx