The hidden questions behind the Heilmeier Questions

This document is a general guide for developing a clear program concept using the ARPA-H Heilmeier questions (HQs). The additional insights and considerations enumerate many overlooked questions underlying the fundamental HQ questions. This guide is not universal or exhaustive in the scope of considerations to address, and candidates should expect additional feedback throughout the application as part of our unique process.

ARPA-(H)eilmeier Questions

1. What are you trying to do? What health problem are you trying to solve?

Provide a banner or vision statement: "Imagine if_____ (fill in the blank)." Explain as you would to a family member or the general public.

- Your end objective should be obvious: a "no-brainer."
    - Think more about the big picture and less about the components.
    - The need for the program must be obvious to all (even if it wasn't before they read it).
    - Avoid jargon. Broadly knowledgeable non-experts must understand and appreciate what it is you propose.
- Focus on the problem, not on the approaches and solutions.
    - Don't unnecessarily limit the program.
    - Though you should have some solutions or approaches in mind, you will challenge the community to bring new ideas to the table.

Considerations: Is the end objective a clear and obvious unmet health-related need? Is the need for the program obvious to all (even if it wasn't before reading about it)? Have you avoided a solution in search of a problem? Have you focused on what the program is attempting to achieve/solve rather than the how (HQ3) or the why (HQ4)? Have you explained the benefit to the US taxpayer if this program is successful? E.g., what "thing" will you deliver that will change health outcomes (a device, drug or method of developing new drugs, or new way of interacting with the healthcare system, etc.)?

2. How does this get done at present? Who does it? What are the limitations of the present approaches?

Define the status quo of today's solutions, the technical state of the art, and why today's solution is not ideal for solving the well-defined problem articulated in HQ1.

- Explain why current approaches can't solve the problem.
- Outline the full range of existing approaches. Show knowledge of the state-of-the-art solution available to the population today.
- Show examples or case studies – don't caricature and be up to date.
- Describe why the current approaches will not meet the program's vision; what technical hurdles stand in the way?
- Use effective reasoning, for example, "The key technical impediments to success are X, and they can't be achieved currently because of Y."
    - Example of poor reasoning: "Nobody is currently doing my cool technology, which is why the problem remains."
- Use this as an opportunity to define the quantitative BASELINE that will serve to build metrics for HQ8

Provide the metrics that quantitatively describe where today's solutions fall short

- Provide data on the effectiveness of today's solutions. These data should define the technical problems addressed by Question 3.
- Clear metrics are critical for driving program phase goals and timelines.

Considerations: Are the state-of-the-art and emerging landscape accurately represented or undersold? Are you sharing up-to-date examples or case studies? Is it clear why current approaches can't solve the problem? Have you explained why a relatively straightforward integration effort involving existing/emerging solutions would be insufficient? Have you fully decomposed the challenges identifying

the fundamental root-cause barrier(s) and/or tradeoff(s)? Have you included input from stakeholders? Have you included key figures, references, and data to support your assessment of the status quo?

3. What is new about your approach? Why do you think you can be successful at this time?

Aim to make a fair and convincing case that moves the concept from "impossible" to "possible." Provide examples of what you will do to solve the problem and how you will do it (and why it's possible now). This is your opportunity to convince ARPA-H that this should be done and now is the time to do it. (You have an excellent opportunity to establish your expertise and share figures and data that offer a glimpse of what could be possible such as new technology X that will enable us to do Y.)

- What are your new insights? Refrain from limiting approaches to specific solutions you might have in mind. Instead, focus on what fresh perspectives or insights you have that suggest the problem may be solvable.
    - Cite recent breakthroughs/capabilities (papers, data, etc.) that provide reason(s) to be optimistic about new possible solutions.
    - Leverage advancements from other communities that have solved analogous problems
- How will your new approach work? Describe how you will implement the new insights you've identified.
    - Outline the framework for implementing your approach, describing how potential areas of focus (these could be parallel research arms that, after two years, combine to deliver the final product) result in program objectives.

Considerations: Have you identified one or more novel approaches clearly showing how the problem could be solved? Do you show what is theoretically possible, what has been practically achieved to-date, and why your novel approach could bridge the gap? Have you made a clear case that your proposed approach(es) is/are "ARPA-Hard" (challenging enough that more traditional funding sources could not support them)? What is the value-added? (essentially, provide clear differentiation from existing or recent solutions). Have you described the fundamental technological or scientific advance that indicates this is possible now when it wasn't ten or even two years ago (to answer the why now question)? Are particular use cases presented (example cases to establish the new paradigm)? What are the new insights that are nucleating your excitement enough to launch a program around them? Are there advancements you could draw on from other communities that have solved analogous problems?

4. Who cares? If we succeed, what difference will it make?

Expound on scenarios of how the program would change the world. Position your approach in the framework of (or in comparison to) existing solutions outlined in HQ2. The objective of the program is NOT to provide a fully elaborated, implemented solution but to sufficiently de-risk it so that relevant stakeholder communities can then fully implement and sustain it. Describe high risk opportunities where the potential solution(s) offer an inflection point, not merely fractional evolution.

Considerations: How will this solution be game changing, and how will we ensure it will "survive in the wild"? Do improvements in metrics justify the investment? Is this a 2% improvement, or a factor of 10x improvement? Who are the target customers, end users, stakeholders, and transition partners? How would they be affected if the program succeeds? What potential transition partners would be interested in in taking the program outcomes to the next level? Is this something they are clamoring for already, or is it a new paradigm that they haven't even considered? If the program does not meet its stated objective, what could still be learned? Are the primary barriers to adoption identified, or does other major uncertainty/challenge need to be addressed before the program concept could be implemented?

5. What are the risks?

Describe the primary reasons the program might fail to achieve its objectives. It is helpful to perform a risk analysis and distill the results to the factors with the highest likelihood and/or severity.
Some of these risks may include:

- **Technical risk**: Will this a require a substantial leap in the state-of-the-art with many technical unknowns? Will subject recruitment be an issue? Will discovering appropriate biomarkers that others have failed to find be a significant problem?
- **Programmatic/management risk**: Will the disparate performers work well together given their very different formalisms and ontologies? Will the results be tested independently and promptly?
- **Transition risk**: Might the American Medical Association be recalcitrant in adopting the new policies? Will the Center for Medicare/Medicaid Services help to make the solution affordable?
- **Tangential risk**: Is the solution amenable to dual use without potential devastating consequences? What technology/intellectual capital protection risks exist (*e.g.,https://grants.nih.gov/policy/foreign-interference.htm*)? Will the solution exacerbate health inequities? (See HQ9)
- **Reputational risk**: Could the government be perceived to be encroaching on civil rights? (See HQ10)

Considerations: Are all major types of risks (scientific, technical, clinical, managerial, translational, economic, environmental) considered? Are risk likelihoods and severities estimated, and are there plans to manage or mitigate those risks?

6. How long will the program take?

Recommended practice is that programs usually last up to ~4-5 years. Typical Programs are Phased (two or three Phases) to break an aggressive program mission into tractable pieces and allow for gates where unsuccessful ideas/teams no longer continue. In relation to HQ8, consider what capabilities performers must achieve at each phase. (Responses to this HQ are often displayed in table format)

Considerations: Is the timeline aggressive but in the realm of the possible? Have you justified the schedule for all expected phases of the program? Note that the program duration can be separated from the Program Manager's tenure. What are the main drivers that might accelerate or decelerate this timeline?

7. How much will the program cost?

Typical ARPA-H investments in a program can range from $20M-$100M over its lifetime. Break down costs by phase and significant task area. Provide brief justification for the expenses. (Responses to this HQ are often displayed in table format)

Considerations: Have you used the targeted number of projects in each potential area and the approximate cost for a project in each area to justify the overall program budget? Is the projected budget for each phase and each significant effort within those phases well-justified and realistic (e.g., order of magnitude)? Do you include a budget for independent evaluation and testing? Do the justifications reflect an understanding of the actual costs? (If applicable) have you included various scenarios reflecting different levels of program scope (full, partial, etc.)?

8. **What are the mid-term and final exams to check for success?**

Metrics should be quantitative (e.g., affects 70% of the population) rather than quantifiable (e.g., we can calculate how much of the population it affects) or qualitative (e.g., it affects a lot of people). They should be ambitious, meaningful, and impactful. (Example: recovery time drops 4x, the cost drops 100x, patient access doubles, infection rate to zero, demonstrate GMP production and IND submission.) For the Phasing outlined in HQ6, what milestones must be met to progress to each phase?

Considerations: Are the milestones realistic, and do they reflect tangible progress towards overall program goals? Are the metrics/success criteria rigorously defined, easily measured, and validated by an independent third party? What are the go/no-go or critical decision points, and have you clarified how you will use them to make decisions at critical phases of the program?

9. **To ensure equitable access to all people, how will cost, accessibility, and user experience be addressed?**

Provide solutions that enable people from all socioeconomic, demographic, situational, and other scenarios access. Address how considerations of cost, user experience, and accessibility will factor into equity.

Considerations: How will relevant stakeholders be engaged with this program? What will maximize and ensure the chances that program outcomes are equitable? How has stakeholder input been integrated into the objectives and outcomes of the program? Is this something that will be affordable? Have you thought of opportunities to improve components, designs, supply chains, etc., to reduce overall cost? Will people have easy geographic access? If there are end-users, will the capability support people with accessibility challenges, such as vision or motor impairments?

10. **How might this program be misperceived or misused (and how can we prevent that from happening)?**

Identify and provide solutions to potential misperceptions and misuses of not only the outcomes of the program, but its concept, formulation, and messaging. Think outside the box and be creative. There are likely misperceptions that you haven't thought of or won't think of. Consider solutions to the unidentified misuses or misperceptions.

Considerations: How might this program be misperceived? By the American public? By other nations? What will reduce this misperception? Will it generate data or other resources that can lead to privacy or cybersecurity risks? Can bad actors exploit the results for nefarious purposes?

ARIEL Example responses to HQs:

The hypothetical program example, **A**udible **R**eal-time **I**nterpretation of **E**xasperated **L**ittle Ones (ARIEL), walks through the Heilmeier Questions and shows how the "big picture" objective and technical goals connect. Note: XX are placeholders. In a real program, concepts use relevant values & units.

1) **What are you trying to do? What health problem are you trying to solve?**
   a) ARIEL seeks to interpret crying in pre-verbal children to reduce household anxiety/distress by quickly connecting with the correct care. When pre-verbal children cry, caregivers can only guess if the cry is motivated by hunger, needing a diaper change, tiredness, sickness, or a missing toy. The parent may not have external symptoms to aid diagnosis if the child is sick. Delays in diagnosing and interpreting distress vocalizations may exacerbate it.
2) **How does this get done at present? Who does it? What are the limitations of the present approaches?**
   a) Trial and Error: Ineffective and may still not identify source of distress; Potential for increased time and costs ($XX spent in the U.S. per year for soothing remedies). Guessing wrong may increase the infant's distress by XX% and lead to XX days of lost work when children are sent home until they recover.
   b) Phone a Friend/Relative: Can provide anecdotal experiences not driven by data; Phone calls are typically not answered.
   c) Take baby to an Expert: It takes too long and can be over an hour in rural communities; potential for misinterpretation (experts misinterpret XX/100 infant cries) and XX hours of lost work per month on average.
   d) Products on the market have been developed with a minimal set of human input and do not represent the full spectrum of infant behaviors (Powell, H. cf. 8F23).
3) **What is new about your approach? Why do you think you can be successful at this time?**
   a) Establish a digital library of sounds, gastrograms, movements, facial expressions, and associated meanings. Estimate recording X hours of audio and video from ~X children, spanning X different stressors across X socioeconomic strata.
   b) Leverage sound/movement/expression library and combinatorial unsupervised/supervised, RNN, etc. ML training approaches to individualize.
   c) Recently developed systems relatively easily parse targeted sounds from ambient noise ("Alexa…"; whale song). Adapting them to child sounds can be reasonably straightforward.
4) **Who cares? If we succeed, what difference will it make?**
   a) Parent/Caregiver: will experience reduced anxiety (a study showed XX% reduction in anxiety) and XX days of recovered work; (be quantitative and ensure that all communities are considered).
   b) Health Care Providers and Insurers: reduce misdiagnoses (cite infant misdiagnoses data, insurance costs, or other data).
   c) Infants: reduced time under distress increases space for cognitive development and reduces chronic effects of stress.
   d) Baby Monitor Companies: Current market size is $XX but will grow to $XX with ARIEL.
5) **What are the risks?**
   a) Technical risks: Unable to attribute meaning to the distressed cries from the baby; unable to sample from enough children to establish sound/movement library.
   b) Clinical risk: need to adapt to each individual child.
   c) Management risk: Team members unable to communicate effectively across disciplines

d) Transition risk: The in-home solution is too expensive to be equitable; unequal access to cloud/WIFI.
e) Unintended use risk: Bad actors hack into systems; Misrepresenting reports to Child social services.
f) Reputational risk: Perception of hidden agendas by the government and/or device manufacturers.

6) How long will it take?

Technical Area	Phase 1 (15 months)	Phase 2 (12 months)	Phase 3 (15 Months)	Total (3 yrs. 6 mos.)
Human Subjects Research	←----------→		N/A	2 yrs. 3 mos.
Computational Methods	←----------→			3 yrs. 6 mos.
Prototype Development	←----------→			3 yrs. 6 mos.

7) How much will it cost?

Technical Area	Phase 1	Phase 2	Phase 3	Total
Human Subjects Research	$6M	$8M	N/A	$14M
Computational Methods	$5M	$7M	$6M	$18M
Prototype Development	$2M	$6M	$8M	$16M
Independent Verification and Validation (IV&V)	$0.5M	$1M	$1M	**$2.5M**
Total	$13.5M	$22M	$15M	**$50.5M**

8) What are the mid-term and final exams to check for success?

Technical Area	Phase 1	Phase 2	Phase 3
Human Subjects Research	• Begin observational study and record from 35% of anticipated needed subjects • Provide initial samples of 3-5 stressors and sound + movements to the Computational team	• Provide full audio & video/distress library to the computational team • Complete observational study recording from 100% of required samples	N/A

	• Re-adjust target sample numbers based on adaptive study design		
Computational Methods	• Identify and adapt existing audio and video detection and isolation tools • Desktop/lab-based demonstration of ability to detect and categorize training set of baby sounds & movements with >70% accuracy (Go/No-Go)	• Desktop/lab-based demonstration of ability to detect and categorize training set of baby sounds and movements with >95% accuracy (Go/No-Go) • Demonstrate detection and categorization of the testing set of baby sounds and movements with >70% accuracy (Go/No-Go)	• In-field/At-home demonstration of ability to associate illness with testing set of baby sounds and movements at >99% accuracy
Prototype Development	• Market research, customer discovery, and end-user surveys to inform product design	• Identify hardware and software needs • Source and design prototype device; establish consumer price point < $50 (Go/No-Go) • Demonstrate bench-top prototype can decipher and categorize baby sounds and movements (Go/No-Go)	• Device prototype evaluated by potential consumers • In-field/At-home demonstration of ability to associate illness with testing set of baby sounds and movements at >99% accuracy with the prototype device

9) To ensure equitable access to all people, how will cost, accessibility, and user experience be addressed?

a) Ensure the subject pool has a statistically significant number of children from different sexes, skin colors, and socioeconomic strata.
b) The user interface supports caregivers who have visual or auditory impairments and provides clear settings to protect privacy.
c) Product design and hardware requirements lead to COGS < $50 with an anticipated retail price of $80 for minimum functionality; premium products may add features and capabilities at higher costs.

10) How might this program be misperceived or misused (and how can we prevent that from happening)?

a) Beginning in Phase 1, engage with parent advocate and pediatrician groups every 3 months.
b) Implement price testing strategies and identify optimal price point for entry in >95% of US homes (comparable to existing baby monitors).
c) During Phase 2 and Phase 3, field test with >100 parents with newborns.
d) Engage in identifying usability, interface, desired features, etc.
e) Implement feedback at three-month sprints.
f) Address intended uses, misperceived uses, usability, cost, etc.
g) Engage external validation and verification experts to evaluate the security.
h) Elicit community surveys to evaluate perceptions of use.