BDHC
B D Hamilton Consulting

AI LESSONS FOR LEADERS FROM ENGINEERS

The Adult Engineer in the AI Room: How COOs and CTOs in chemicals, natural gas utilities, oil and gas, and pharma and biotech make AI decisions they can defend.



Ben Hamilton, PhD

PhD Chemical Engineer, ex-Worley, ex-PwC

THE UNNAMED PROBLEM

In 2020, inside PwC, the teams I was building AI products with crashed into a problem that did not have a name. It was not that the engineers weren't up to the job. They were brilliant, and the work kept failing for reasons outside their control.

In consulting and audit environments, tech breaks quietly. Teams move fast when they are dialed in, but most teams that try to move fast fail. The failures can be large, yet they stay invisible to decision makers because they land in business “intangibles” that seem impossible to measure.

Move those risks into heavier industries—chemicals, natural gas utilities, oil and gas, pharma—and the same failures become physical disruptions, supply chain interruptions, regulatory scrutiny, loss of trust, and reproducibility crises. The promise of AI pulls one way, the operating models of these industries pull the other, and good decision making can feel like falling behind the headlines.

In my experience, the ways of working that make most businesses successful and the ways of working that make AI products successful are oil and water. Sometimes that gap is worth navigating, and sometimes it is worth avoiding. I have better language and tools for it now. That is what this is.

WHY THIS EXISTS, AND WHO IT IS FOR



I have been talking leaders out of AI since 2018. I am an engineer and systems modeler who has spent a career designing industrial, laboratory and software systems, and I have little interest in building technology that doesn’t work. I want to solve hard problems at meaningful scale, with real-world constraints, in ways that are measurable and ethical.

Whether your operating environment is physical or informational, the vulnerability is the same: probabilistic software, used by, near, or instead of probabilistic humans, placed on top of brittle processes. That exposes you to a breadth of unknowns, and very few organizations are navigating them well. Leaders are surrounded by vendors speaking in hype-driven buzzwords, but they lack an objective translator who can bridge operations and technology, audit the processes and risks, evaluate data maturity, and tell you whether an initiative protects capital or burns it.

If you are weighing a data or AI related decision right now with a knot in your stomach—whether it is a board demanding an AI strategy, or demanding AI with no context, or a $500k pilot stuck in pilot hell, or a vendor that can’t evaluate their own work—pause. I’ve got your back. You are probably not mismanaging this, and your teams are probably not failing. You all may be burning out, though.

When these initiatives struggle, visibly or invisibly, the pressure lands on the team. These are people who will do what they are asked, partly because AI is exciting and partly because nobody wants to tell a manager no, even when they know what to say no to. I've seen teams chase metrics that map to nothing real, build systems from too little information about the workflow, and work inside incentives that undermine their real goals. Capital burns, your best people get exhausted, and no measurable value arrives.

I help identify the operational outcomes that protect your capital and your people. That can mean examining strategy and organization, or the decisions and workflows of individual operators on the floor, as focused or as broad as the problem needs. I stay vendor and software neutral, so the only thing I optimize for is your outcome.

WHAT WE WILL COVER

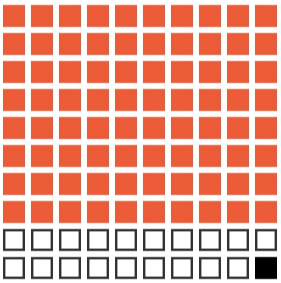
- Where AI initiatives fail, and why model performance accounts for only a fraction of those outcomes.
- Six recurring conversations I have with executives, and what they all have in common.
- Evaluating which technology can hold up to your goals and concerns, financial and governance included.
- Calculating AI ROI ahead of time when most folks think it is impossible.
- Why data maturity sets an innovation ceiling, and why you don't have to raise it all at once.
- When to use off-ramps, when to stick with it, and how to identify misalignment a year or more in advance.

WHERE AI ACTUALLY FAILS

I talk people out of AI largely because the hype is so strong right now. Most of what is being peddled out there isn't helping people, and it isn't helping businesses meet their objectives. I've been working in this space since 2018, and honestly, the buzz was around back then too. Data scientists were needed everywhere, machine learning engineering was a baby discipline, and 80% of data science initiatives failed.

You'd think that number would get better with time, not worse.

AI fails to deliver value about 95% of the time. I've evaluated around two hundred AI cases and seen nine figures of waste.



HOW THE 95% GO WRONG

AI is rarely the reason AI fails.

80% fail because of organization, process, and people.

19% fail because of the technology and data enabling AI.

1% fail because the AI itself underperforms.

The failure modes take 18 to 60 months to see, and I've seen longer. They look productive, but deliver no value. They often make a team or a technology look like a bad investment, when something else was in the way, or absent. Fantastic ideas can be set up for failure years before they reach that end.

A good team will do what you tell them to. A good vendor will sell what they believe in. But if the org can't handle what's required to assess and integrate the parts, it's wiser to find where that bottleneck is, work out whether it can be addressed, and say "we may not be able to do this."

THE SIX CONVERSATIONS

Leaders I talk with typically start at one of six initial conversations:

"Build me my AI idea ASAP."	"The CEO says it's time to get AI."	"I am pretty sure I need it."
"I don't trust AI."	"I don't see the value."	"We've already got or are building AI."

None of these contain the important question. *What does your business actually need to achieve, and what will it take to get you there?*

Answering it properly takes your expertise, and sometimes a translator who can thread your needs through the capabilities of this technology wave. A robust process picks the right tool for the job: advanced machine learning, traditional automation, another intervention entirely, or no new technology at all. Whatever the path, the 80/19/1 reality holds. Most of the risk sits in the human, organizational, and process factors that get ignored while everyone is transfixed by the technology.

Do all six end up in the same place? Without support, mostly yes. Some folks get lucky: their approach works without their knowing which conditions set them up for success. That's a bad gamble, given how long it takes to find out things are wrong.

WHAT THE MARKET IS PITCHING

Here is what the market is actually selling you:

- AI is inevitable.
- AI is taking half of all jobs. Aimed at leaders, and rolled back when the public backlash was severe.
- A reality your brain automatically sees as a great thing. Works on me too.
- The same reality, framed as something your competitors are doing, or your business will fail because it's old or behind.

Relentless marketing says AI is the next big thing and says nothing about how to make it valuable in real operations. Sound engineering ideas like retrieval augmented generation (RAG), knowledge graphs, digital

twins, and agents have been turned into buzzwords. The only sure thing is that a lot of folks worry they are missing out, or that somebody else is outmaneuvering them. My worry is different: the real value is there, and the marketing has people looking in the wrong places.

The most intelligent-seeming chatbots are sophisticated products designed to maximize engagement, not to provide unbiased intelligence. Because the interface feels natural, it is easy to mistake a polished prototype for an enterprise-ready system. In heavy industry and regulated environments, the gap between a working demo and an operational tool is substantial.

Try an easy experiment. Tell your chatbot to give you curt advice with no flattery and no sympathy, only concise detail and transparency about its own biases. I find that useful. If the products were designed that way, far fewer people would buy them. The marketing is strong enough that even folks who know how these systems work get pulled in. That's not a character flaw. Words carrying novelty, authority, or identity slip past our mental filters, so we never stop to think.

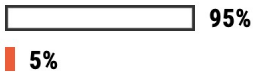
It looks so good out of the box. Nothing in AI has ever been this polished, believable, and compelling. It sure looks like it works. It feels like it works. Helps us be busy. Lets us cut the headcount.

BEN'S RULE FOR SETTING AI EXPECTATIONS

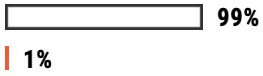
- **Traditional Machine Learning:** When a system looks 90% complete, it is about 10% done. Getting to production means scaling time and resources for edge cases and data drift.



- **Generative AI:** When a prototype looks 95% complete, it is 5% done. The polished interface hides the engineering still needed for governance, retrieval accuracy, and error handling.



- **AI Products:** 99% and 1%.



- **Agents:** They need high data maturity and standardized operational definitions. Built on probabilistic components without rigid data structures underneath, they have no viable path to operational value.
- **Autonomous Systems:** Many operational workflows need neither machine learning nor generative AI. Let business constraints and engineering principles pick the technology.
- **Baseline Workflows:** If a process doesn't work without AI, understand its baseline mechanics first. Instrumenting the process, whether physical, computational, or human, lays the foundation for any software layer and moves even ambiguous ways of working toward process engineering.

There are plenty of examples where these tradeoffs are still worth it. Making them work in your context takes attention to things we've never had to think about.

Take a document assistant built on retrieval augmented generation. The demo shows a language model gliding over your file shares. In production it meets unindexed, inconsistent, decaying business records,

response quality drops, and adoption stalls. That isn't the model failing. It is your data maturity ceiling showing up, and getting the balance of performance right takes work. There is no perfect.

Test at lower levels of complexity first, with error analysis, human feedback loops, and security constraints in place. Then leadership can fund the data work with eyes open, or take a clean off-ramp with confidence. Each step up in scale brings a new set of edge cases, and you meet them from a stronger position.

"You sound like you are trying to talk me out of AI."

AN EAST COAST COO

Exactly. Let's talk about real business challenges and how I can help.

CHOOSING WHAT IS WORTH BUILDING

"Don't build stuff that isn't obviously valuable enough to be worth working through the unknowns."

STRATEGY DOWN, PROBLEM UP

When the board demands an "AI strategy," leadership teams often respond by collecting a dozen competing pilot proposals, then quietly wonder how to separate real business value from meeting-room novelty.

Valid AI opportunities emerge from two directions: top-down organizational strategy, or bottom-up operational pain.

Top-down evaluation treats AI as one possible intervention serving organizational goals, subordinate to your data and technology strategy. If AI itself is in your strategy, that is often a sign it will end up aiming at the wrong objective. Get the strategy clear, the priorities right, and the problems valuable enough, and AI can be designed in where it fits. No more, no less.

This framing also helps when the right answer is something other than AI. Robotics, RPA, decision optimization, and more work amazingly well when they fit the problem, hype or no hype. What you need may be far simpler, and more impactful.

Bottom-up evaluation starts from acute pain felt on the ground. These problems sit closer to real value, so it is easier to judge whether they are worth working through the unknowns. They are some of the most fun to solve, and no less complex than top-down ones. Here too, rigorous problem solving often shows AI is the wrong tool.

The big risk in sourcing ideas is copying what other people do because they're doing it, or dragging in a use case because "it would be cool if." AI is never the right reason for AI, any more than a shovel is the right reason for a shovel. Don't ask AI people what to build. Source it from real problems.

ROI IS A DISTRIBUTION, NOT A NUMBER

Once an initiative comes from operational reality, its financial case is the second filter. For leaders used to physical capital projects, this feels unfamiliar. Technical unknowns, organizational friction, and behavior shifts show up faster and broader than in traditional engineering work.

Traditional financial models assume deterministic outputs and fixed returns. Probabilistic software doesn't behave that way, which is why you will hear that AI returns can't be calculated. They can.

Evaluate value as a probability distribution of outcomes rather than a single number. Start broad: does the potential upside justify exploring, with the downside explicitly priced in? Here is an illustration from a high-value, low-frequency operating environment where elite operator time is the main constraint:

- **The Floor Case:** You invest enough capital to test whether an approach yields meaningful return—for instance, a $400,000 investment targeting an ROI range between 50% and 400% by capturing one additional high-value outcome. The objective is determining whether the baseline potential makes the attempt worthwhile to validate the approach before scaling.
- **The Worst Case:** You take the investment from the floor case and treat it as a total loss (-100% ROI). A well-designed experiment can result in a -100% ROI and still be valuable if it prevents a multi-million-dollar zombie project. Most failed AI initiatives bleed capital because an initial $400k pilot quietly expands over twelve to twenty-four months with no additional validation.
- **Scaling Up, current reality.** What good looks like in your current reality: the same team, now at full capacity with the technology in place. Stopping here can already be significantly impactful.
- **Best case, the capacity breakthrough.** Fundamental process redesign. New capabilities you couldn't achieve before. Compete in previously impossible segments. Requires rigorous risk management through every phase.

$400,000 → $80,000

Managing Risk by using off-ramps to cap downside exposure from $400,000 to $80,000.

In most AI projects, exposure piles up because teams wait too long to halt failing efforts.

Seen as a distribution, the value of fast learning becomes obvious. A controlled failure at -100% ROI after two months is often an operational win, because it stops the capital burn and preserves team bandwidth. Total exposure depends on how quickly leadership enforces an off-ramp.

Design experiments to surface failure modes early, and an expensive mistake becomes a disciplined, high-velocity lesson.

DATA SETS THE CEILING, WHICH YOU MUST RAISE

You need regular access to data owned by finance, maintenance, or another operating group to make an AI project useful. Months pass in alignment meetings, security reviews, and access requests while the initiative sits dormant. That is when the quiet doubt sets in: Why is retrieving our own data turning into an organizational battle? Am I failing to lead this?

Or your technical team proposes data platforming, cleaning, or pipeline work, and your first reaction is skepticism. You remember the last data warehouse that ran over budget, took two years, and delivered little operational value. Are we about to sink money into another endless infrastructure project?

It's OK that your data is siloed across plants, offices and labs, in systems that were never designed to feed AI systems. It's OK if you tried this before and it didn't work out. It's OK if you tried data science but had to let them go because the data was in rough shape. It's OK not to know the state of your data. All of that is normal. Skepticism is healthy and justified.

Data maturity is a prerequisite for AI maturity. Commercial tools are limited by the data they were built on and the tasks they were built for. Custom and fine-tuned models are limited by the quality and structure of your own data. And data that supports human decisions needs different handling from data that feeds machine learning.

Past data initiatives often failed because they tried to platform and clean all enterprise data at once, with no anchor in specific workflows. The market now pitches AI as a way to skip data preparation entirely. In reality, skipping data readiness stalls AI, and boiling the ocean wastes capital.

As models get more complex, or move toward agents, extra data layers become necessary. Every business term you take for granted needs a standardized, explicit definition. Agents can be a valid engineering choice, but they need high data maturity, and for now they only work in narrow use cases.

You do not need a multi-year platform overhaul to move forward. Pick a single operational workflow, identify the minimal data it needs, agree on shared definitions for those inputs, and leave the broader silos alone for now.

Design governance and adoption together. How data flows in, how operators work with inputs and outputs, and how errors get flagged are one requirement, not three. Iterating the human workflow alongside the technical pipeline keeps you out of the trap of systems that look busy in development and deliver little in operations.

ITERATION AND MEASUREMENT

Work backward from the maturity level where your eventual need sits to where your organization is today. Do not skip stages; the technology will let you try, but it won't work. Identify what the next stage requires and iterate into it. A common trap is iterating on the tech long after the team, the organization, or the operating approach is what needs the update.

Make the first steps small enough that errors show up sooner. You accelerate organizational learning by rolling smaller dice more often and correcting in calibrated steps. It isn't glamorous, and nobody will celebrate your sensible, quiet rollout. The best execution looks incredibly boring.

Measurement goes hand in hand with iteration, and traditional project metrics fail here. Out-of-the-box vendor metrics like tokens processed, queries answered, and activity logs measure software activity, not business value.

Useful measurement works less like a pass-fail scorecard and more like focusing binoculars. Early readings have wide error bars and low resolution. The point of early iteration is information gain, steadily reducing uncertainty about performance, edge cases, and the human experience. If an initiative can't show tightening signals of growing operational value every six weeks or so, it will languish. The metrics are a mirror, a crucible, and an opportunity.

The best execution looks incredibly boring.

THE 80 PERCENT, ORGANIZATION AND PROCESS AND PEOPLE



A technical team delivers a working system on schedule, and weeks pass with zero meaningful adoption. Operators quietly bypass the tool. It is easy to wonder whether you hired the wrong people, or whether your team is resisting change.

Usually it is neither. Workflows get built by people who don't do the daily work. Software adds friction instead of relief, so operators route around it to protect their throughput. KPIs reward speed while probabilistic tools demand careful review, so the organization accidentally rewards rubber-stamping errors. Accountable

leaders who are spread too thin never engage enough to see what changed, and end up asking for things that undermine their own success.

WHAT GETS IN THE WAY

- **Ignoring user research.** Teams skip it because they assume technical expertise is enough, and build tools that miss the human experience success requires.
- **Politics over process.** Teams protect individuals and circle the wagons to keep the peace instead of following the needs of the work.
- **Misaligned incentives.** The exact same person can be highly productive or badly stuck depending on whether the environment gives them the right triggers.
- **Diffusion of responsibility.** People feel less personal obligation in groups. Everyone agrees on a course, leaders delegate, and nobody owns the outcome. Add AI to the mix and it happens more often.
- **Mismanaged Centers of Excellence.** Left to choose their own projects with no guidance from the top or the floor, they drain capital. They languish as gatekeepers, or chase hype without an eye on business value.
- **Context-blind leadership.** Leaders reinforce abstract values instead of providing the concrete context and constraints their teams need to make better decisions.

WHAT TO DO INSTEAD

- Treat direct user research as a non-negotiable engineering requirement, even for systems that look heavily automated.
- Use tight, iterative cycles to expose institutional friction early, instead of big, rigid change programs that tend to fail.
- Design incentives and governance around how people actually behave, not how the org chart says they should.
- Define and assign unambiguous ownership for every initiative, so accountability doesn't evaporate into the structure.
- Ground the CoE in real relationships with business units and subject matter experts. I designed ours around enabling, not gatekeeping.
- Give operational context early, before stakeholders see the work, so decisions align with the business.

Many people I have worked with are better than they think they are. Many are better than their orgs will ever appreciate. These gaps are bridgeable.

GOVERNANCE, AND THE WORKFLOW UNDERNEATH IT

Governance has to be thought through AHEAD OF TIME. Choose models, infrastructure, and integrations without that diligence and you build a system that solves the wrong problem. Governance isn't optional if you want the system to succeed. It is only optional in the literal sense: these systems are so accessible that you can deploy AI far faster than you can think through its implications. When AI makes decisions humans can't sufficiently audit and control, you add variability.

The biggest mistake I see is forgetting that a probabilistic system cannot guarantee deterministic outputs. No math gets around that. Without a human in the path whose workflow gives them final authority, safety relies on hope rather than design.

"Human in the loop" has become a market term, often abused into compliance theater. Real governance looks at how human responsibilities shift under pressure. When operators move from creating work to reviewing it, fatigue sets in. Pushing human decisions through probabilistic models raises error rates over time, and expertise quietly erodes because it is no longer doing the primary work.

Most failure stories have no party responsible for the impact of the change. A leader accepting responsibility doesn't count. Delegating it doesn't count. It works when workflows are designed carefully: the decision points, how people interact with them, and the user experience all the way up to the leader who signs off. Generic systems fail even when they look effective.

Which is the real point: The map is not the workflow. The workflow is the sequence of decisions people actually make, the information they have when they make them, and what happens to the work when an error slips through. When governance is designed hand-in-hand with the operator's daily reality, safety and adoption become the same operational outcome.

KNOWING WHEN TO FOLD

AI risks stay hidden because human brains don't have the biology to detect them. Interfaces present polished, fluent outputs, statistical thinking is riddled with bias, and leaders project deterministic reliability onto models. We are wired to anthropomorphize anything that talks back, reading intent, agency, or intelligence into statistical output.

AI is not a colleague or an autonomous agent, not even the ones called agents. It is code, statistics, and computing architecture. When a decision gets made because AI seems human, that is our brain failing to tell the difference, not the truth.

Here is my contentious point: What if AI use cases should fail to bring value 95% of the time? Experiment wisely, off-ramp sooner, pick more valuable problems, and count every idea that doesn't meet the cut-off, and 95% is a great number. Some ideas may come back around later, but never for their own sake. We might as well get it up to 99. The best AI is the one you don't have to build, because a simpler solution does the job better.

In engineering, every decision has an actual engineering decision underneath it. When technology scales as fast as it does with AI, those decisions get made whether or not anybody makes them on purpose. Which is why I preferred investments to fail gracefully and quickly rather than limp along, burning resources.

HUMAN ENGINEERING, FACE FIRST

At the risk of sounding philosophical, AI engineering is human engineering. Technology won't replace humans, partly because our internal systems are different, and that difference is exactly where the two can complement each other.

Much of the industry has positioned these tools to extract value from people rather than build it, and the public trust lost along the way will set the field back. That is a shame, because when AI is the right tool for

the job, it offers real breakthroughs. AI doesn't make decisions or eliminate jobs. Leaders do. Operational changes, workforce transitions, and system redesigns are human choices, and the responsibility for their impact rests with leadership.

One of the clearest signals of organizational health is how leaders talk about their teams. Calling a workforce reduction "AI efficiency" is an abdication of responsibility to the people whose lives are intertwined with those organizations. The leaders I work best with evaluate, support, and transition their people responsibly through technological shifts.

Sound execution means clear-headed analysis, and technology deployed in service of human capabilities and concrete business outcomes.

THE PATH FORWARD

There are a few ways to describe what I do. I'm the same person no matter which you ask for.

Technology consulting, not AI consulting. AI is my wedge. It's everybody's wedge. I'm a 20-year multi-discipline systems thinker, who helps you connect your business and technology to value.

AI Products and Tools: I solve technology for my clients from 0 to 1, meaning my experience runs from strategy and conception through build and adoption, to value and maintenance.

Process Consulting: A collaborative approach that helps organizations get better at solving their own challenges.

Process Engineering: AI systems, software and business systems are all process engineering. This includes workflows, human-technology interfaces, and org structures.

WHAT YOU GET

- Expertise that meets your business at an angle, across strategy, commercial, regulatory, operations, and technology.
- Better decisions, faster, by threading your team's and your SMEs' experience together with mine.
- Unviable builds killed before capital is committed, avoiding multi-month waste.
- Regulated and sensitive data barriers reframed and solved as explicit engineering problems.
- Vague AI asks become scoped, fundable and risk-managed programs.
- AI that gets built, integrated, and adopted around your operational reality, and runs independently with your team.
- Hidden technical, governance, and organizational failure modes caught before production deployment.
- A steady hand and clear communicator from initial strategy to operational maturity.
- Help realigning existing investments and debottlenecking processes that seem impossible to measure.

Ready to talk? Connect with Ben **HERE**

WHEN TO ENGAGE

- Your AI strategy requires objective evaluation before committing capital to unverified projects.
- Your KPIs are generic, intangible, or the path to measurable value is unclear.
- You've bought or implemented AI and have growing concerns about its governance, adoption, and efficacy.
- Your data infrastructure lacks the maturity required for advanced capabilities.
- You're managing massive technological shifts and don't know what you don't know.
- Friction, resistance, or slow communication at any level is stalling a great team or a great idea.

WHAT I DON'T DO

I engage as a partner to the leader who hires me, not as a spare pair of hands, though I'll deploy where needed and roll up my sleeves with your team. I don't push a specific technology or vendor, AI included. I help you find the right fit, I have a reason for every recommendation, and I won't suggest work that has no value. I'm here for hard problems with real constraints.

WHY PEOPLE STAY

"It is helpful to have a data scientist who speaks software engineer, AND real scientist, AND farmer."

"I didn't realize at first that you were learning my business until I saw it come together."

"They will automatically feel a reduction in stress and anxiety. It is great to have an expert on board who knows AI inside and out and, in one meeting, could clearly identify what is slowing us down and getting in the way of our efficiency."

"You are the missing piece of the puzzle."

My goals are twofold: One, help you meet your objectives. Two, help you build capabilities so your organization can solve more challenges in the future. Engagements are value priced, so I can match feasibility to the value of your challenges. Initial engagement meetings are flat fee.

Network. I have referral partnerships and contacts with fantastic builders. We find what's right for where you are at, and tap my contacts when appropriate. They all know I don't push vendors, and respect me for the boundary.

If you got to the end, you may have a challenge worth a call.

BOOK 30 MINUTES WITH ME HERE



ABOUT THE AUTHOR

Ben Hamilton earned a degree in chemical engineering from Purdue University, followed by a PhD from the University of Minnesota. His early work was in pharma, nanomaterials, and chemical processes, where failures had massive costs and visceral consequences. By 2009 he was automating his own job, building simulations and models, and analyzing things everybody took for granted. He spent several years as a process engineer consultant at the height of the fracking boom.

In 2017 he combined that process engineering and simulation expertise with data science and AI. He has been consulting in AI since 2018, and is roughly a couple hundred AI cases deep.

His record includes low nine-figure impact across the energy work during the fracking boom while at Worley, and at PwC, eight-figure mistakes avoided, eight-figure opportunities identified that would otherwise have been cut, and cultural transformation reaching more than 6,000 employees. His methods for discovering AI impact come from creatively navigating constellations of risks for problems worth pursuing, linking strong decision making frameworks, and a healthy skepticism of what the market says AI can do.

Ben operates as a consultant through BDHC, established in 2019, helping leaders solve for and measure business outcomes when AI is both so promising and so hyped.

He lives in Seattle, trains Krav Maga, and defends his title as Favorite Uncle. He teaches chemistry and robotics to kids, and lately soap chemistry to adults.

Email: ben@bdhamilton-consulting.com
LinkedIn: linkedin.com/in/bdhamilton
Website: www.bdhamilton-consulting.com
Book a call: calendly.com/ben-bdhamilton-consulting/30min

