

Who and What Drives Algorithm Development: Ethnographic Study of AI Start-up Organizational Formation

RODNEY SAPPINGTON, Ph.D.

Acesio Inc.

LAIMA SERKSNYTE, Ph.D.

Acesio Inc.

The focus of this paper is to investigate deep learning algorithm development in an early stage start-up in which edges of knowledge formation and organizational formation were unsettled and contested. We use a debate by anthropologists Clifford Geertz and Claude Levi-Strauss to examine these contested computational forms of knowledge through a contemporary lens. We set out to explore these epistemological edges as they shift over time and as they have real practical implications in how expertise and people are valued as useful or non-useful, integrated or rejected by the practice of deep learning algorithm R&D. We discuss the nuances of epistemic silences and acknowledgments of domain knowledge and universalizing machine learning knowledge in an organization that was rapidly attempting to develop algorithms for diagnostic insights. We conclude with reflections on how an AI-Inflected Ethnography perspective may emerge from both, data science and anthropology perspectives together, and what such a perspective may imply for a future of AI organizational formation, for the people who build algorithms and for a certain kind of research labor that AI inflection suggests.

Keywords: AI, Deep Learning, Algorithm R&D, Epistemology, Domain Knowledge

SETTING THE SCENE

We are in a 3,000 square foot office space and from a small board room we hear “that’s domain knowledge—we’ll get subject matter experts for that!” Team members debate over hiring radiologists and hiring expertise outside of algorithm development. They are puzzling over the usefulness of radiologist knowledge in developing deep learning algorithms that are to serve radiologists in their diagnostic interpretations. We have arrived at an early stage deep learning startup in Silicon Valley near a strip of park land where cyclists and joggers stream by an area that once fell into neglect and was reborn into a corridor of fusion restaurants and tech companies.

We find ourselves among decades of tech venture capital infusion set against university and residential growth. You can smell chlorine from Olympic sized pools and spas, coconut-crusted shrimp from a high-end bistro and musty ash from recent wild fires. Flocks of feral parrots sound off in the trees high above, escapees of nearby ranch-style living rooms. We have arrived at what social network theorist Mark Granovetter has termed an “innovation cluster” (Ferrary & Granovetter 2009) of artificial intelligence (AI) or what Andrew Ng has termed the “the new electricity” (Ng 2018). The term refers to algorithms that could be described as “rocket engine[s]” in which “huge amounts of data” can be processed lifting the rocket ship of machine intelligence to new heights (Garling 2015). Jeff Dean, the head of AI

at Google assesses Deep Learning algorithms as keys to AutoML, machines that learn to learn and are foundational for early detection of a wide range of diseases (Dean 2018). Such start-up companies that Peter Diamandis describes as “smaller” and “nimble” have evolved out of a linear industrial era into an “exponential” era, a period of market disruption, unpredictability and 100x algorithmic growth (Koulopoulos 2018).

Based on our real experience of such a start-up, the concepts and possibilities of our case study of Deep Learning algorithm development and new organizational formation emerge out of this exponential petri dish.

STRUCTURE OF PAPER

This paper explores original research set within deep learning algorithm development in an early stage start-up organization, from 2014-2015. At the time I (Rodney Sappington) served as a Senior Data Scientist focused on developing algorithms for the early detection of disease. Together with my collaborator we analyze features of deep learning algorithm development which include observations and interactions with strategic partners, clinical leaders and radiologists. We also refer to experiences of highly regarded team members with whom I was involved and had the pleasure in part to lead in algorithm and clinical development. Not the least we have developed our perspective out of references across the social and behavioral sciences and discussions with colleagues outside the walls of algorithm development.

The structure of this paper includes conceptual and case-study analyses. We contextualize our research by way of briefly introducing a passionate debate between Clifford Geertz and Claude Levi-Strauss on the fear and embrace of computational forms in ethnographic practice. This debate is not a gentle or dated conceptual tug of war between renowned anthropologists but instead a contemporary lens through which we can investigate contemporary algorithm development and messy problems of building algorithms with medical diagnostic capabilities. We briefly explore the appearance of deep learning algorithms in Silicon Valley. We turn to early stage start-up features and move to the complexities of a case of a health insurance company that offers large amounts of data and a problem for algorithm development with a twist. We go inside the organization and examine hiring practice and the role of domain knowledge in a machine learning start-up with a scene at a company retreat. We explore the problem of positionality of being a data scientist and anthropologist in the field of applied machine learning. We attempt to open up how the diagnostic patient is being conceptualized. In conclusion we provide observations on epistemological tensions in algorithm development which we term *AI-inflected* ethnography. We are not suggesting a methodology but instead pointing to a way of inhabiting these tensions and silences at the core of algorithm development, an approach that opens up a view into how organizations and organizational members get constituted and sometimes unravel.

FEAR-EMBRACE OF COMPUTATIONAL FORMS

There could not have been two more different scholars studying human behavior and culture than Clifford Geertz and Claude Levi-Strauss. For Geertz field research was immersive, forged in dust, blood and side-bets surrounding Balinese cock-fights. For Levi-

Strauss field research was forged as a product of field work in conceptions of binary order and kinship system. For them a certain kind of intelligence and mind of the ethnographer was at stake, it was almost at once human-driven and machine/system-driven:

Society is, by itself and as a whole, a very large machine for establishing communication on many different levels (Levi-Strauss 1953).

Lévi-Strauss has made for himself...an infernal culture machine. It annuls history, reduces sentiment to a shadow of the intellect (Geertz 1973).

On one side Geertz viewed ethnography as a practice of interpreting human speech and gesture in the wild and intentionality in everyday life, which he called “deep play.” On another side Levi-Strauss viewed ethnography as a scientific practice of interpreting universal codes, totems and patterns across society, it was “structural.” The gist here was a tension: a type of human perception and cognition emerging *in* local everyday life versus a type of human perception and cognition emerging as universal patterns *across* everyday life. For Levi-Strauss information science held a central place in creating and interpreting culture which implied both human and non-human forms. For Geertz information science suggested an infernal (hellish) culture machine.

There was another kind of legacy that was epistemological. It drove their passion and still fuels passions today in machine learning. As anthropologists and data scientists will still largely live and work in this legacy, categories of “hard” and “soft” knowledge, universality and particularity, probabilities and possibilities of quantified judgment and human intuitive judgment, structured and unstructured data and embodied and cognitive forms of intelligence. I take the Geertz-Levi-Strauss debate as a struggle for thinking how we build, imagine and fashion algorithms for human benefit and machine automation.

DEEP LEARNING ALGORITHMS EMERGENCE IN SILICON VALLEY CONTEXT

Andrew Ng and others in 2012 built high level features using deep learning to recognize and classify cat videos at a 70% improvement over previous “state-of-the-art” networks (Quoc et al 2012). Using computational power, 16,000 computer processors, the deep learning network was presented with 10 million digital images found in YouTube videos. This was a breakthrough and supplied proof that certain deep neural networks could perform (automatically learn without human hand-coding) across complex image sets. This was the beginning of the successful use of convolutional neural networks (CNNs). Two years later overutilization of medical imaging in healthcare made radiology ripe as a testing ground to apply some of the lessons learned from 2012. New early stage organizations began to take shape to productize these findings. The context was Silicon Valley in which a line-up of similar innovations had set an aspirational and venture capital stage for deep learning algorithm development. Expectations were high. From Apple, Uber, Lyft, Google, Airbnb, Salesforce, Tesla and Twitter to name a few, today nearly 30 technology companies are so-called “unicorns” or near “unicorn” status totaling close to \$140B in value (Glasner 2018). It has also brought together venture funds. Venture funded machine learning start-ups has recently almost defined the San Francisco-Silicon Valley region in terms of company

valuations. As the global economy is expected to grow by 3.5% GDP, venture backed AI startups have had an expected growth rate of over 40% by 2020 with a U.S. market valuation of AI by 2035 of \$8.3T (Faggella 2018). Along with this growth came warnings that we “must be thoroughly prepared—intellectually, technologically, politically, ethically, socially—to address the challenges that arise as artificial intelligence becomes more integrated in our lives” (Faggella 2018).

The new start-ups are often described fondly in the industry as “Moonshots.” Singularity Ventures founded by Ray Kurzweil also terms them “exponential startups.” These are organizations that claim to hyper-scale across person, transaction and global impact. Founders within them are typically referred to as super smart. They exemplify that the social good of machine learning goes beyond a mere transactional machine or consumer recommendation system or ad placement. As the perspective goes these are exceptional people building exponential machine learning products with exceptional resourcefulness. What they know and how they apply what they know has become a global phenomenon that others try to emulate.

This smaller-nimble organizational type has been described by Ferrary and Granovetter as one “node” in “a durable” assembly of organizations that has an almost magical capacity to “anticipate, learn and innovate in order to react to major internal or external changes” (2009). This is a lot to ask of people building the new global electricity with industry expectations to touch every human life on the planet.

DEEP LEARNING EARLY STAGE START-UP FEATURES

How did these expectations measure against our real-life start-up experience? Features of our early stage start-up were less than magical and included product discovery instead of product development, immediate press coverage and AI hype, fear of replacing radiologists by robots, extreme leadership pains, and intense VC oversight and intervention at the expense of a vulnerable organizational culture. It comprised of a mix of possibility, struggles over knowledge to be brought into the organization and algorithm architecting and iteration. Trial and error were essential ingredients to gaining replicable results and to rapidly building an early stage start-up.

In daily operations, the team’s approach to capturing disease was not at first to understand it but to start with a particular, by identifying the features of lungs, their shapes, edges, anomalies. Our team had to first segment thousands of lungs before we could begin to achieve any results with our algorithms in identifying lung nodules. This meant using chest x-ray and chest CT’s pixel data and annotations from publicly available data sets that were by no means perfect data and required preprocessing and extensive labeling. The lungs are a vascular world onto themselves. Millions of veins spread out into the lungs and different types of scar tissue can be present that obscure cancerous nodules until late stage becomes incurable. In building our algorithms we were working with convolutional neural networks (CNNs) that have the ability to auto discriminate image features.

This was an exciting time to work at an early AI start-up. Deploying and testing CNNs was a creative endeavor in 2014 when these algorithms were not in wide application for medical imaging analysis. It was a time in which venture capital investors were smitten with deep learning algorithms and needed only a good team and a good idea to trigger an investment. It was a time in which the horizon of what was algorithmically possible in

medicine was at an inflection point but the practical application and proof of good performing algorithms were sometimes daunting to demonstrate. Additionally, this excitement came with the price of reducing human/patient complexity to the purity of an all-powerful algorithm that could be generalized across medical contexts.

Ideas and conflicts abounded. They were worked through when we took ‘one-on-one’ walks. We walked along grassy walk ways puzzling through who to bring in as consultants or full-time employees (FTEs). We white-boarded approaches. As I walked with the lead scientist we often considered bringing in oncologists, radiologists, primary care physicians. Social scientists were seen as “not useful.” The social sciences and ethnographic knowledge were a hinderance to successful algorithm development at this time. Clinical expertise was considered but walled off, kept as consultants, advisors, reduced to domain knowledge. In other words, the highly skilled area of radiology could not travel well beyond radiology but data science could travel and exceed radiology workflows, image interpretation, disease classification, sub specialties and the human. What we faced during this period was how to formulate problems and how to formulate a diverse team who held particular forms of knowledge appropriate to the problems we were trying to solve. It was difficult and slippery discussions. No one seemed to have a magic-bullet answer.

When faced with a lack of diversity of knowledge around algorithm development Jeff Dean, Head of AI at Google has stated:

I am personally not worried about an AI apocalypse” but “I *am* concerned about the lack of diversity in the AI research community and in computer science more generally” (Dean 2016).

Not only he was concerned about encouraging people from different backgrounds to build algorithms, he was thoughtful that certain forms of thinking may not get into algorithm development. Experts from the Google Brain Residency program, which would have been a feeder for such diversity recruitment were composed of “physicists, mathematicians, biologists, neuroscientists, electrical engineers, as well as computer scientists” (Dean 2016). These were largely STEM practitioners. This range of diversity did not include unexpected perspectives. Diversity appeared narrow, bounded.

The fear-embrace of computational forms could be viewed as an epistemological tension between those whose knowledge contributed to algorithm development and those whose experience and knowledge was viewed as consultative. Such consultative knowledge was typically referred to as “domain expertise” and could be reduced to a kind of consulting artifact. As indicated in Dean’s statement domain expertise may not have gone completely unacknowledged, it was called for but not followed through with as evidenced in Dean’s listing. A STEM defined in this way could screen out social sciences, physicians, policy experts, artists, ethicists, community members and patients to name a few. Such screening out came in the form of particular/local knowledge that was perceived as not algorithmically scalable across industries. Could ethics scale across industries? Could a patient’s experience of navigating and overcoming a deadly cancer and a fractured healthcare system scale? We are not questioning Google’s idea of inclusiveness or diversity. We are pausing on what/who gets persistently divided up as contributory and held up as core knowledge in algorithm development in an organizational context.

When it comes to marginal, diverse or unexpected perspectives not held in high regard in algorithm development, two words come to mind – *problem formation*. The kinds of

problems that get identified and privileged for algorithm development are shaped by the kinds of people who are brought together to identify and attempt to solve those problems. Problem formation is as valuable as problem solving. A red light is as valuable as a green light for a specific algorithm project. When diversity of perspectives, experiences and background is a slogan-only proposition, acute problems simply are invisible to machine learning innovation or worse, they are seen as exciting problems that have positive social value when they instead have potential negative societal consequence. Different kinds of problems drive different outside engagement practices.

PROBLEM FORMATION COMING THROUGH THE DOOR

Machine learning problem formation could come through the door from strategic partners and data providers. Problems were not always defined from behind the organizational door among team members. Locating and defining a problem that fit the capabilities of algorithm development was sometimes called “product-market fit” but in terms of our case it was also an effort to locate large data sets and then allow the problem to emerge from the data. An attractive offering of large data sets could supersede a more sober problem formation process. The team around the table and their backgrounds and training often determined which problem got selected that could set the organization down a developmental road for months and years.

An Example

It was a bright brisk northern California day, the sun bouncing off pavement at a nearby corporate square. It couldn't be more gorgeous. For months we had been negotiating the terms of a partnership with a large health insurer that promised near-term revenue for us. The health insurer team arrived outside the front of our building and discussed things before collectively announcing themselves and walking in. They greeted us in sharply dressed suits and strong handshakes. We all entered a meeting room and settled around a glass table with a white-board of diagrams of medical imaging archiving, workflows and model layers. I turned the white board around and they began by introductions and then launching into the billions of transactions they do each day. One of their team members a man in his mid 50s hair combed straight back who appeared to be a 1970's version of a suburban executive, call him J., began describing their value in terms of transactional data and the possible extent of a strategic partnership with them, “as you know, we have all the data on the patient journey, how the patient is treated and if they go to X hospital to Y pharmacy, rehabilitation center, pharmacy - you know, we have the whole thing.” He punctuated his brief description by “we have more patient data than we even know.” It was billions of transactions. The project as they described it, was for us to build algorithms using this massive transactional data set.

“We want to eliminate unnecessary [insurance] audits, get rid of them all together if we can,” he said.

“You mean a percentage of the audits you conduct?” I responded.

“Yes, they're a waste of time.”

Their machine learning lead engineer, a man in his mid-thirties in a plaid shirt began to lay out some ideas around predictors and unsupervised learning strategies that could work together to assist in this direction. The goal was to identify the probability of medical fraud and reduce

unnecessary insurance audits. They were looking for precision in what was known as “fraud detection.”

Algorithm development was to provide a means to rank medical practices in terms of probability of committing fraud not in terms of actually committing fraud. On one level such algorithms could save private practices the back-breaking process of unnecessary insurance audits that could cause mountains of paperwork, anxiety, and administrative time. They wanted our algorithm development to share in a moral victory for physicians and physician practices, we would be the good guys, less audits less onerous oversight less unnecessary phone calls and less legal expenses for the private medical practice. It was a hero’s problem and we could be the heroes to solve this problem.

To get the problem in focus I paraphrased it for our meeting to make sure I understood. “You want us to help you build predictive models based on transactional and temporal data that would save millions of dollars of wasted effort in wrongful audits of potentially thousands of private practices.”

“Not wrongful, but yes to help save millions in unnecessary audits.” He said, they would invest and provide all the data we needed.

Around the table I could feel a kind of moral victory flag being raised. But something was not quite right.

I asked if their mission of reducing audits was all they were intending with our algorithms. He said it was “hard to determine future uses” and technology was “moving so fast” but this was absolutely their “main use case.”

“If we are building algorithms to assist you in identifying *less* unnecessary audits couldn’t these same algorithms be repurposed to help you identify *more* audit opportunities? You want us to help you reduce audits and I understand this, but couldn’t you just as easily increase your audits with our technology? Couldn’t you become an audit powerhouse in some way?

“That would not be good for business” he said and then defensively mentioned “we’re looking for the right team, it’s a great opportunity.” I felt our CEO shift in his chair. We needed the data.

Then J. said something more interesting. “In reality”, we “rarely commit unnecessary audits, our approach to audits even if they do occur are over years.” He took a drink of water as if to refuel and then said, “for example, we never audit a practice twice in the same year and most practices never get audited at all.”

My own experience was different. Coming from a private surgical practice background I had been part of exactly such unnecessary insurance audits and their impact on staff. These effects were fresh on my mind. In fact, my practice was audited twice in the same year by this same insurer with no upcoding or wrong doing of any kind found. I was the one who actually experienced this back-breaking administrative work first hand. I experienced late hours, certified letters, operation reports and patient records reassembled daily based on changing requests from the insurance auditing department. How could he have known that I as a data scientist could possibly have experienced this same insurance auditing process. I kindly responded.

“I think that may be inaccurate, my practice was audited twice by you last year and we came away with [our] claims in order.”

He quickly shot back. “That’s very unusual, we don’t conduct audits this way. The name of your practice?”

The point was that they did indeed operate this way and audited more than once in the same year and unnecessarily. From their perspective such algorithm development was a win-win saving time and money for their insurance company and saving human distress and labor for private medical practices across the country. However, “future uses” were to be determined.

Digging deeper, algorithms that were designed to decrease audits could be weaponized to increase audits across millions of medical practices. Physicians could learn quickly to avoid complex patients that might have presented a risk for a billing error or they might experience another consequence of the ongoing threat of audit, and bail out of independent practice altogether and become an employee of the local hospital leaving billing responsibility and legal exposure to the hospital. The ongoing threat of medical audit could reshape networks of private medical practice ownership. No doubt medical fraud has been a key challenge in U.S. healthcare with the justice department in July 2018 announcing 1.3 billion in fraudulent claims across doctors and treatment facilities. A large number of these were for over-prescription of opioids and false billings. On the other hand, delinquency aside, the impact of unnecessary and aggressive insurance audits known as “fraud detection” could collapse a resource-strapped medical practice, drive physicians from owning their own practices, encourage consolidation of medical practices by large health systems and breed a culture of reimbursement fear. There was irony in an early stage machine learning organization asked to take on such a project that could accelerate the destruction of the very kind of organizational ethos it holds so dear: entrepreneurship.

In the current climate of insurance audits it was not the evidence of fraud but the evidence of a mistake that could trigger an audit. A single mistake could trigger a multi-year audit over hundreds or thousands of patient encounters. In my previous experience a misspelled word in an operation report could trigger a process that would lead to a claw-back of hundreds of thousands of dollars. It could cause crushing legal fees, employee burn out and an ever-present anxiety of the next audit always around the corner. Just the threats of audits could be the quickest means for driving out independent ownership of medical practices and the quickest way of controlling (reducing) the complexity of patients that a physician accepts. The more complex the patient the more likelihood of a mistake, no matter how small or administratively mundane.

In our meeting the question was who/what would set the criteria for such fraudulent probabilities? Would criteria be set by this specific insurer? By the insurance industry, by Medicare/Medicaid or those holding federal office? Features could be identified and built into algorithms and customized over time. What defined an outlier billing event could shift, and such outliers could be categorized as suspicious. Such “suspicious” billings events could tilt towards criminalizing practices.

Thus, problem formation coming through the door could hold both, human benefit and harm. We discovered the negative impact of fraud detection at this time not because we were skeptical of health insurance companies but because specific experience was at hand that gave dimensionality to the problem they offered and scope to its downstream cascading possibilities. This expertise was domain specific but was central in making a decision for the entire organization. However, herein laid a problem regarding domain expertise which most typically was disregarded in machine learning start-ups.

DOMAIN EXPERTISE AND HIRING PRACTICES

When we considered problem formation it was best to examine expertise as it was being recruited into the organization that had the potential to shape such problems. Our recruitment process had cognitive, technical, behavioral, and problem-solving dimensions. All the tests on potential new hires were offered in a ‘one-on-one’ context with a third person taking notes. Criteria for ranking candidates was subjective and based on who conducted the test and on the testing criteria although the criteria itself was viewed as objective with the same problems and same questions asked of each candidate. Additionally, a percentage of candidate testing was partially conducted over the phone or Skype.

We recruited and tested for what we called T-shaped skills. We were hiring for depth of skills, which represented the vertical part of the T and ability to collaborate in cross-functional teams, which represented horizontal part of the T. Data scientists/engineers were all supposed to possess T-shaped skills.

Nevertheless, the overall hiring process ran against this T-shaped approach often identifying skills only useful for coding, and setting aside other skills as consultative, domain specific or non-T-shaped. This blind spot in our hiring was rarely acknowledged although we urgently needed skills focused on disease classification, high mortality cancer and a complex healthcare system, all outside of a typical coding experience and seen as “other” skills. These other skills were not T-shaped but rather singular, particular, walled-in domain knowledge. Broadly, we needed T-shaped radiologists, business leaders, nurses, patient advocates, IT managers and operational managers, we needed a T-shaped organization. Instead we hired data scientists. What this did to us and to our organization was to miss and not encourage people in roles who could translate forms of clinical and technical knowledge across product, business and algorithm development, we missed a translational T-shaped organizational culture.

How did this “T-shaped” hiring process that missed good cross-domain knowledge got expressed in the process of algorithm development?

Domain Knowledge: Universality and Particularity

We arrived at a company retreat by a northern California lakeside surrounded by redwood and pine forests. Swimming pool and cluster of cottages were set among a small gravel parking lot and hiking trails that ascended the hillsides. The retreat was called in somewhat a haste and before we knew, we were all gathered together to discover improved teamwork and sharing of ideas. All our activities were constructed to give everyone a voice, to ask questions and make suggestions on how we could improve the company. But a strange pattern emerged. As machine learning engineers and data scientists held more and more air time, business development, operations, a few clinical consultants had less and less time to make suggestions. Throughout the retreat the problem increased with less and less time allotted for cross-functional expertise outside of coding to offer opinions. The emerging organization was getting locked into its own limited self-perception.

On the last day of the retreat with only 20 mins left a colleague raised his hand and questioned why so few suggestions were asked or offered outside of engineering. The room got quiet. The CEO said “that’s not true, everyone has been included.” Then my colleague went further. I will call him Paul.

“You’ve been soliciting suggestions from everyone in your mind, but in this room it has been with only maybe half the company.”

The CEO expressed disbelief: “no one’s been excluded!” And then, “it’s everyone’s responsibility to speak up.”

The actuality was different. When someone spoke up about problems or solutions outside of coding, the Q&A would move quickly on to the next person. The function of not having enough time to be democratic during a company retreat was not the point. It was a function of certain forms of knowledge that were recognized and other forms of knowledge left unrecognized or invisible. A group company retreat was in fact a fractured retreat of what one colleague said “ideas to be considered” and “ideas that were awesome,” and “creative” and “productive”. In this context the ideas that received the greatest attention concerned a narrow focus on machine learning algorithms, medicine as strictly a data problem at scale, machine learning expertise as universal knowledge, and team development and cohesiveness around these issues.

What was relegated to the “water cooler” conversations were ideas regarding lack of product direction, building and recruiting a clinical team with cross-functional responsibilities, leadership issues, a diverse organization at scale, work/life balance and low team morale.

Let us briefly unpack these ideas as they were expressed. The “awesome” ideas regarded machine learning as scalable, central to organizational development, around which all other forms of knowledge revolved. Outside this sphere clinical and business knowledge were consultative or being driven by algorithm development. Machine learning knowledge was the epistemological glue that bounded all other operational forms of technical and organizational mastery.

The second-tier ideas revolved around the organization and its people with their knowledge and experience. Disregarded clinical, cultural, retrospective feedback on mistakes and product direction could in fact act as a key to the emotional and professional bonds that reached inside and outside work-life balance. We believe the seeds of innovation were contained here, in this second tier of ideas that could not be well heard. The organization went on retreats and participated in team exercises but their contributions were not well captured nor well understood. The idea of what was universal and what was particular (domain-specific) knowledge were typically translated into concepts that were framed as “useful” or “not useful.”

Read or Experience

The following conversation between a team member and a radiologist, and later the reporting of their conversation to the team may serve to expand on this universalizing-particularizing tension in our machine learning start-up.

“How do you see that [spiculation]?” - a colleague who we will call Tim.

“I compare it [nodule] to others I’ve seen” - the radiologist who we will call Don.

“It’s what you have seen or what you know from what you’ve read?”

“Years [of] experience” - Don explains.

In our team meeting comprised of product and engineering this conversation was translated by Tim as a certain kind of data-driven opportunity for algorithm development.

“You guys, we can detect nodules by edge characteristics. Radiologists have diagnostics in their head and I think we can get the same diagnostics from imaging.” Tim’s comment became clearer and got reinforced in the following.

“So, we hire radiologists to annotate and we build, annotate and build, right” - says the CEO.

“Not that simple but that’s the idea, we bring in domain experts.”

“We plug them in or we hire them full time?”

“I’m not sure we’ll know what to do with them full time, the best would be consultants.” Tim speculates.

Domain expertise has been established here as a contributory source of expertise. It provided a certain kind of fuel to algorithm development but was not valued as central to such development. It could be read and learned without having to be experienced. Radiology expertise was reduced to an “industrial designer: everything involved in a situation, any kind of material, must be calibrated for its contribution as evidence” stated anthropologist Marilyn Strathern (2004). The radiologist as a domain expert was seen as delivering a snap shot of knowledge and yet providing knowledge gained over years of experience of disease identification and classification. Except, that experience was disregarded as “useful.” This artifact had to be used as evidence that could be codified into algorithm development. Such expertise was not seen as evolving and aspirational, it was seen “useful” only as evidence for creating clinical labels for algorithms. The team’s excitement of using radiologists as consultants and not knowing what “to do with them full time” had reduced radiologists’ value into a kind of automating piece work.

Contemporary radiology was often seen as a medical science of piece work. The radiologist has been well acquainted with holding a consultative role across medical disciplines with the widespread use of high-speed broadband and three decades of teleradiology. Medical images travel instantly, as radiologist interpretation travels remotely to hospitals and clinics, new imaging technologies continue to develop. Radiologists have been participants in a somewhat similar experience of reducing knowledge and their own knowledge to a commodity. Speed, distribution of radiology report and harsh efficiency have long attended upon radiologists. At first sight, it did seem to be a perfect collusion between algorithm development and teleradiology practice. However, the piece-working of radiology practice as a discipline into teleradiology practice has been different from its use as domain knowledge in a machine learning start up. The difference was radiologist’s autonomy and agency to practice their craft that has always been held in high regard and to develop and perfect their skills has been an ongoing commitment. In an algorithm development context, the opportunity to perfect radiologist skills went one direction, toward algorithm building and scale, radiologist-consultants were plugged in and the race to gain new diagnostic insight has been transformed into a race to produce diagnostic evidence useful for algorithm iteration.

One of radiologists’ key concerns at the turn of the 20th century was how to enable physicians to quickly reason with diagnostic insight while accompanying patients at bedside. One of the key medical concerns today has a twist: how radiologist’s insight is to collaborate with/aid machine intelligence to make life saving diagnostic judgments across billions of data points in an instant.

ANALYTICAL EYE/I

In this section we wish to describe shifting role boundaries in a startup. As a Senior Data Scientist my role included tasks across disease classification, deep learning research, business development and data wrangling for model training. I often occupied a ‘wear-all-hats’ position in which there were more tasks than people and more roles than could be filled.

Trained both as data scientist and anthropologist the borders of my skills have never been settled. This meant that my ways of thinking and valuing people’s insights were not fixed in terms of algorithm development. On one level psychically (a debate in my head) and outwardly (a debate with colleagues and product agendas), a kind of slippery battle waged across domains of knowledge that were valued among one group and not valued among another group. The value I placed on knowledge and expertise was always open for revision depending on the tribe I was working with. This led to occupying various roles even when my title stated a very bounded one. Carrying out various tasks, often contradictory ones within certain roles has been a professional tension that has persisted. For example, I often went from data scientist to product and partnership development taking back from partnership meetings client suggestions that were not recognized until only much later or too late. I caught myself speaking ethnographically instead of with the brevity of an engineer with a definable problem and outcome. Observational thinking was occasionally valued, but often not, and sometimes even scorned. The anthropologist’s seeing with an eye towards levels of knowledge and a data scientist that must get “shit done” did not always rest easy together. Ethnographically stepping back and technically stepping into highly productive algorithm development was exciting but it could also be exhausting. One had to be elastic and thick skinned.

Authenticity was a sticky matter. I had to have constantly revisionary set of understandings on how data science, anthropology and machine learning specifically worked together and for whom. I lived and worked among revisionary sets of ideas that were never wholly laid out and never fully given over to my colleagues. Being authentic as a trained data scientist, deep learning researcher and anthropologist was not, however, a negative affair, it was one that had no user manual and no clear-cut boundaries. It did require summoning creative parts in oneself. It was a skeptical stance towards algorithms, and an enthusiastic stance towards their possibility.

Role boundaries and formal positions within our start-up could crash into product sprints, team meetings, and patients and physicians we were supposed to serve. This was our company’s way of operating professionally. It was a specific way of working and acting upon knowledge that was cherished and knowledge that was left opaque. This was my direct experience of the organizing potentiality and growth of data science *thinking*, and not just its *application*.

THE CONCEPT OF THE “PATIENT”

One other role, the role of the patient, also eluded the makeup of our early start-up.

Patients as abstract beneficiaries, radiologists as domain experts, hospital administrators as users shaped our evolving and tenuous company culture. An embodied sense of what a patient was or who such a patient could be was often abstracted and would appear lost to busy organizational members. However, the patients that were to be served and the

radiologists whose diagnostic skills were to be enhanced by algorithms were not really lost but instead were not shared among team members. ‘One-on-one’ walks could not get at shared understandings and would veer off into my or your medical story. There were many “patients” many “radiologists” many ideas of how algorithms would “enhance” the diagnostic encounter. The patient was always elusive and suggested a different kind of ethnography in order to render up these clinical-technical nuances that were embedded in such a machine learning organization, algorithm development and ideals and organizational members under intense pressure to build, scale and commercialize.

The problem of getting the patient in view was mixed with the problem of getting our models in view for productization. We remained far off from an actual product. We struggled with what kind of patients would benefit from our work, were these younger/older patients, smokers/non-smokers, male/female, those who only could afford sophisticated algorithms applied to their chest CTs or MRIs, were they local or in the developing world, were they near-term or years away from benefiting from our work?

A team member brought this confusion into relief when he read from a patient letter he had received. He was the one practicing MD on the team. The patient’s letter was heartfelt and delivered to his home and he brought it in that day to bring through the door an embodied sense of a patient who had undergone both, good and poor diagnostic experience. He had been his patient. This is a paraphrase of the letter.

Dear _____

I wish to thank you for what you have done for me and my family. I’m not sure if you remember me but I was the difficult one who kept asking questions and you were the one doctor I could rely on during my time that tried to answer them. My wife said you were in and out most of the time but I felt you were at my side. I’m not sure how you found the tumor when others missed it, I guess I don’t know how doctors cannot see things that can kill you and someone like you can see it. I really don’t understand medicine but I know you were there and helped us and I’m grateful to you.

I’m delivering this to your home because I wanted to make that effort to come to you as you must have come to me at my bed[side]. I think you saw me as a person not as another patient, I want to believe this and will hold onto it.

Again I and my family wish you much happiness and success in the future and again thank you for your professionalism and care.

Sincerely,

This was delivered in a company-wide meeting. It was an intervention of sorts in an attempt to ground our efforts in an ethics of patient care and an embodied picture of an actual patient who was spared by good diagnostic care and who was grateful and alive to attest to their care. It was also a way to open up a discussion on the consequences of diagnostic error which our algorithms were to correct and reduce. Questions were asked about this patient but then we moved abruptly into an investor meeting. This quick move away from the reality of this letter was indicative of knowledge and experience that *could not* be taken in or absorbed. A few months later doctor colleague brought up this incident with a sense of shock that any notion of a patient just floated and could not be grounded in our push to build diagnostic algorithms that would help to save such person’s lives. Between the

care of algorithm development and diagnostic care of patients a huge gulf existed. The perceptual limitations of algorithm builders were well on display.

Upon reflection the awkwardness of the team's responses or non-responses to this patient's letter was of a different order. It conjured up many patients, many diseases, and many possible projects that brought out the ambition in team members to assist doctors in avoiding missing a critical diagnosis. The letter paradoxically, paralyzed and mobilized the team. Our team was not trained to take on and manage patient suffering, they were not trained to hold threats of mortality or the threat that mortality may come anytime around the corner by a missed diagnosis of an aneurysm or a missed adenocarcinoma (most common form of lung cancer). The team was not emotionally capable to both, take on the threat of diagnostic error, or to accept the overwhelming gratitude that comes from having another chance at life. The work of algorithm development and the work of giving more years to a person's life were at professional and experiential odds.

As a data scientist the teams I have worked with have fallen along a spectrum of well-integrated to chaotic. Our team was well-integrated at times and fell apart into chaos at others. Had I been an ethnographer coming in and out of the organization these moments may have been missed and I would have come away with a completely different picture of team dynamics. For example, after this letter was delivered we floated an idea of a "patient committee" inclusive of patients who had experienced lung cancer and who perhaps had a missed or under diagnosis. Everyone had a different idea of what a patient was, who a patient was and what kind of patients we could recruit. The topic of patient insights being integrated into algorithm development eventually went by the wayside and were reduced to domain knowledge, sent to the periphery of the organization and put on the shelf as patient consultants, when needed. The conception of the patient was marginalized even when it was emotionally charged and required for a deeper understanding of doctor-patient interaction, diagnostic product workflow development and understanding of a medical professional's insight and error. Even when we clearly needed to hear the voices of patients as part of our R&D, somehow we could not accommodate those voices.

Our operations manager put it well: "we all have been patients but we can't imagine their needs when we're here. It's like we stop thinking of ourselves as whole people, here we are only parts of our selves. It's strange, my father died of lung cancer, you have extensive experience in the field but the team has a hard time mapping these experiences to their work." I asked her what she was really getting at.

"It's hard to build what we're building [algorithms] on pain and suffering of others."

I reminded her that we were building algorithms to avoid such suffering through enhancing radiologist perception, to save lives. The patient's letter was a "success story." She shook her head as if to say, *that* was only part of the story when we were outside the organization, on retreat, but not accepted inside the organization where the "real" work got done. Emotions and patients with all their suffering and pain were messy, algorithms were to protect us from that messiness.

Revisiting Geertz and Levi-Strauss

One of the key concerns of Geertz and Levi-Strauss was how knowledge traveled, was taken up and applied to reveal everyday forms of life including human potential and its limits. The hesitancy-embrace of the computational could be seen in this light as a tension in realizing

human potential to see, categorize and celebrate every day cultural forms and cultural others inside ourselves and outside in organizational life. It was on one level a debate on the uses of domain expertise of their era versus the uses of universalizing cybernetic systems. On another level it suggested a contemporary tension of universalizing and particularizing forms of knowledge that must live side by side in practice, and in our case, inside an early stage machine learning start-up. Computerized data is not only organized into training data for machine learning algorithms but is *organizing* of resources and people with all their complexities. Such data also organizes *possibilities* between computerized agents that “think” and “act” in the space between medical problem formation, forms of knowledge and hiring, and algorithm testing and validation. This universalizing-particularizing epistemic adjacency if embodied and used within contemporary organizations would expand upon and not foreclose upon possibilities of algorithm development.

Levi-Strauss and Geertz were not struggling over the present but rather struggling over a certain kind of everyday future in which human possibilities were circumscribed and discoverable by thinking machines. For Levi-Strauss social life was composed of “universal laws”, for Geertz social life was brought forth by the freedom of mind of the ethnographer that had to negotiate intelligent agents that had potentiality to act, feel and displace the knowledge of the ethnographer. Levi-Strauss looked for universal patterns in everyday encounters, Geertz looked for hidden meaning in everyday encounters that was always being found, taken-up again and contested by local inhabitants. As often as they disagreed they both were looking for patterns and universal attributes grounded in small hard-to-access local forms of human life. They shared a focus on making visible possibilities of humanly discoverable evidence in field research, which held different challenges and different opportunities. How local knowledge was shared and how it moved among and between local inhabitants and among ethnographers was crucial to their notion of other forms of thinking and behavior. At a base level they were interested in the ethnographic mind in the context of automation and universalizing computational systems, sometimes posed as a threat, sometimes an embrace, always rubbing up against each other but never losing sight of a human who feels.

CONCLUSION: PUZZLING THROUGH EPISTEMIC EDGES IN ALGORITHM R&D

We have examined dimensions of problem formation in data science and the splits in certain forms of expertise and knowledge that could have practical implications for the shaping of an early stage machine learning start-up. When we thought about the procedures for defining the borders and scope of our case study we became increasingly aware of the shifting role boundaries, problems in-process, and people and their expertise being marginalized and yet continuing to push ahead with algorithm development.

A case study that was dynamic with multiple lines of flight took us back again and again to puzzling through the human messiness of data science knowledge and what got put outside or was made as other in machine learning knowledge today and in the foreseeable future. Machine learning knowledge did not travel alone without being enmeshed in people, organizations, bodies, data and patients, and caught up in certain denials and acknowledgments of human pain and suffering. In other words, we acknowledged the

observable tensions and looked for ways to advocate for silences lurking underneath the obvious.

Methodologically, capturing algorithms required daily ethnographic note taking, mundane but consistent participation, and an interpretative stance similar to Geertz's that allowed one to depict the speech, gestures and uncertainties that came from everyday organizational challenges. We suggest, such analysis drew from a knowledge in data science at least an appreciation of how data was not just wrangled, handled or acquired but was *organizing*. To work with computerized data (in our case healthcare data and medical image data) occasioned being organized by its constraints, possibilities and by strategic partnerships as we have seen with our health insurance company focused on fraud detection. Such consistent participation in the life of an early stage start-up allowed us to gain a larger sense of the textures of data and human patterns similar to the spirit of Levi-Strauss that took us out of the particular and into a generalizing conception through algorithm development.

As we have seen, machine learning could be considered universal viewed with cross-industry application irrespective of specific expertise that has shaped those industries. In our case, a universalizing dimension was applied to algorithms, a domain specific boundedness was applied to the content creators of those industries. Thus, one of the negative outcomes we encountered was epistemic splitting, instead of epistemic inclusiveness, parsing knowledge into plug-ins and universalizing machine learning concepts above other forms of knowledge that weakened organization and compromised quality and flow of ideas that ran through algorithm development. Splitting knowledge into bits also split people and eroded organizational bonds that could have forged professional curiosity across disciplines. Domain expertise suggested a depth of knowledge in a particular field, however it might have also suggested an ongoing attempt to be pollinated by and extended into other areas of expertise. It had a potential to be a truly T-shaped enterprise.

A Note on AI-Inflected Ethnography

The idea of an AI-*Inflected* ethnography emerged at this time. It came together for us as a way to bring qualitative analysis to silences and to the downstream impact of algorithm development at the moment of conception and problem formation.

We believe an AI-*Inflected* ethnography has a potential to focus on the silences, intentionalities, gaps, aspirations and conceptions that may be in one moment accepted and in another moment forgotten or rendered invisible. An AI-*Inflected* ethnography we suggest, is not about what has been built but what has not yet been built and the reasoning or emotions that go into and are fought over to arrive at the product or algorithmic problem worthy of engineering time and worthy of gaining sometimes expensive data resources.

When we think about AI-*Inflected* ethnography we should not conjure pictures of methodological reasoning. Perhaps such reasoning will come over time and over further case studies but here we are marking active algorithm development today and its implications on people in a start-up that struggled and often failed to inhabit different points of view. An "*inflected*" ethnography focused on AI development does not offer up suggestions for research tools, recording procedures or discussion on synchronous (real-time) or asynchronous (non-real-time) of research data capture and technique in the field. Instead we have chosen to examine dimensions of problem formation in data science and

epistemic splits in certain forms of expertise that have had practical implications and anxieties within an early stage machine learning start-up.

This type of analysis we believe is best located in moments when we can tease out the collaborative opportunities and imagination in algorithm development in an everyday when diversity of thought may be up for grabs and when algorithm problem formation may be held open. This form of analysis is unstable and in some ways an outcome of role murkiness between data scientist and ethnographer. But such analytical instability comes as a benefit. It is at these formative moments when certain epistemic events can be made visible through gaps in idea generation.

One of the strengths of *AI-Inflected* ethnography is paying close attention not only to the content of knowledge, but also to the processes of knowledge development focused on shifting edges of algorithm R&D. These edges of organizational knowledge and practice are not simply learned, known and then applied ideas but are locally contested and puzzled through. Deciphering epistemic borders suggests that everyone in an early stage start-up is engaged on one level or another with processes of the formation, co-creation and development of algorithmic knowledge and outcome even when these outcomes are very uncertain and distant. The process of building algorithms paradoxically engages and silences organizational members. The epistemological glue that represents machine learning knowledge does not draw together all forms of knowledge as domain knowledge or as a binary of usefulness/non-usefulness. Instead this glue provides a kind of organizational coherence as much as it limits the flow of forms of knowledge and fashions types of algorithms and types of organizations. These inflection points in AI development are tensions and realities not to fend off and avoid but to identify and transform.

It is important to keep in mind that what characterizes an *AI-Inflected* ethnography is not consultative attendings but an organizational embeddedness of an ethnographer/researcher that helps capture gaps and slippages within an organizational environment operating as a kind of disjointed body pulling together its internal fortitude to take on and build algorithms that can be integrated into software and into people's lives. The odds of early stage start-up success are daunting, the odds of their failure are well over 90%. Could these odds be improved upon with a true T-shaped organization and a true diversity of perspective?

Where Do We Go From Here? - What we gained access to was a deeper image of an organizational group of people trying to genuinely integrate a patient's 'thank you' letter into their thinking but unable to seize upon its message of gratitude and transform this message into actionable algorithm development. We gained access to the potential consequences of radiologist's insight as consultative input and domain knowledge. We gained a sense that the generalizing and particularizing aspects of machine learning were open and unsettled. There were forms of knowledge and emotional life that algorithm development resisted. We believe there may have been other forms of knowledge and diversity of knowledge that has not yet been organizationally imagined to produce very different algorithms for very different outcomes. We do not yet know how these forms will appear in the future but we can as researchers prepare for their emergence by laying the groundwork for the conceptualizing of more malleable and inclusive algorithms.

One of the most disembodied technologies in development today requires embodied and embedded research and engagement. AI calls for us as researchers to not only step back

into the social but to step into the daily grind of AI's silences and epistemic threads that are constantly being shredded and mended to transform organizational coherence.

If we use algorithm development to screen out the pain and suffering of vulnerable people like patients, instead of finding ways to integrate their journeys, then we may also find ever-narrowing algorithm development populating or even taking over our everyday lives.

It is an extensive kind of labor to evaluate, question, chart, chronicle how a start-up and AI development together forge forms of thinking and acting across algorithm, people, product and clients. It seems to be a worthy endeavor. However, if this kind of research needs to get done who is going to do it? Who will listen and make use of it? What kind of organization will it produce? Most importantly, what kind of commitment will it take to produce it?

Rodney Sappington Ph.D. is a data scientist and ethnographer who leads machine learning strategy and product development. His research focuses on the intersections of knowledge, human-machine relations and behavior, and societal outcomes from algorithm development. Rodney is CEO and Founder of Acesio Inc. rodney@acesio.org

Laima Serksnyte Ph.D. has 15+ years of experience in advanced research, consulting and coaching in the areas of executive leadership and organizational psychodynamics. Her research focuses on psycho-social, behavioral and societal levels of analysis in areas of AI development, healthcare, education and organizational member advancement. Laima is Head of Behavioral and Organizational Research at Acesio Inc. laima@acesio.org

NOTES

Acknowledgements – Thank you to the reviewers and curators of EPIC, especially to Dawn Nafus for her continued insights and feedback. A big thank you to colleagues in applied deep learning and data science for careful consideration of the concepts in this paper.

1. A “unicorn” company is a start-up that reaches \$1B in revenue.
2. “Moonshots” are typically defined as ambitious and aspirational projects and companies reaching to develop bold and future-oriented products and services.
3. As a cautionary note, our educational and psychological systems have to keep up with AI projected development otherwise fall behind and in disuse, creating an elite group or risk a population whose knowledge is woefully behind, not relevant, or worse considered a danger to AI development. Bladerunner and many other dystopian societal images come to mind.
4. Claims (insurer) transactional data can have multiple uses in machine learning. Claims data is standardized and medically coded and covers a wide area of the patient's journey. For example, prescription and behavioral trends can be captured across pain medications (opioids) and cholesterol lowering drugs (statins) and this same data can also be used to track physician practice claims, types of claims and time points when outlier claims could indicate probability for fraud.
5. Spiculated margins of a lung nodule are uneven edges that can indicate a higher risk of cancer. “Most nodules that are not cancer have very smooth or rounded margins or look like several rounded nodules together (also called “lobulated”). See Lung Cancer Alliances explanation “Understanding Lung Nodules: A Guide for the Patient.” https://lungcanceralliance.org/wpcontent/uploads/2017/09/Understanding_Lung_Nodules_Brochure_dig.pdf

REFERENCES CITED

- Dean, Jeff
2018 Deep Learning to Build Intelligent Systems by Jeff Dean. AI NEXTCon Silicon Valley 2018, April 10-13, Santa Clara.
<https://www.youtube.com/watch?v=4skC-kGCYRA>
- Dean, Jeff
2016 The head of Google's Brain team is more worried about the lack of diversity in artificial intelligence than an AI apocalypse. *Recode*. August 13, 2016.
<https://www.recode.net/2016/8/13/12467506/google-brain-jeff-dean-ama-reddit-artificial-intelligence-robot-takeover>
- Faggella, Daniel
2018 Valuing the Artificial Intelligence Market, Graphs and Predictions. *Techemergence*. Updated September 16, 2018
<https://www.techemergence.com/valuing-the-artificial-intelligence-market-graphs-and-predictions/>
- Ferrary, Michel and Mark Granovetter
2009 The role of venture capital firms in Silicon Valley's complex innovation network. In *Economy and Society*. 38:2, 326-359.
- Geertz, Clifford
1973. *The interpretation of cultures*. New York: Basic Books.
- Garling, Caleb
2015 Andrew Ng: Why 'Deep Learning' Is a Mandate for Humans, Not Just Machines. *Wired*. May 5, 2015.
<https://www.wired.com/brandlab/2015/05/andrew-ng-deep-learning-mandate-humans-not-just-machines>
- Glasner, Joanna
2018 The Unicorns Behind San Francisco's Burdensome Startup Success. *Crunchbase News*. April 18, 2018
<https://news.crunchbase.com/news/unicorns-behind-san-franciscos-burdensome-startup-success/>
- Koulopoulos, Thomas
2018 According to Peter Diamandis and Ray Kurzweil, These Are the Most Dangerous and Disruptive Ideas. *Inc*. January 19, 2018
<https://www.inc.com/thomas-koulopoulos/according-to-peter-diamandis-ray-kurtzweil-this-is-how-you-build-future.html>
- Levi-Strauss, Claude
1953 *An Appraisal of Anthropology Today*. Eds., Sol Tax and others. Chicago, IL: University of Chicago Press.
- Ng, Andrew
2018 AI is the new electricity. *Startup Grind*. March 6, 2018.
<https://www.youtube.com/watch?v=fgbBtnCvcDI>
- Quoc V. Le, Marc Aurelio Ranzato, Andrew Ng, Jeff Dean et al
2012 Building High-level Features Using Large Scale Unsupervised Learning
<https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/38115.pdf>
- Ruiz, Rebecca R.
2017 U.S. Charges 412, Including Doctors, in \$1.3 Billion Health Fraud. *New York Times*. July 13, 2017
<https://www.nytimes.com/2017/07/13/us/politics/health-care-fraud.html>
- Strathern, Marilyn
2004 The Whole Person And Its Artifacts. *Annual Review of Anthropology*. 2004. 33: 1-19