Getting Real Value from a Labour Market AI Tool: Opportunities, Best Practices & Critical Success Factors

Evan W. Steeg, PhD & Kathryn Wood for Eastern Ontario Leadership Council
December 14, 2020

0. Executive Summary

The Eastern Ontario Leadership Council is undertaking an ambitious project aimed ultimately at the creation of an AI-powered data analytics and forecasting capability that would assist the region's governments, economic development, business and community leaders in making better-informed decisions about labour market priorities and related initiatives to advance the regional economy.

Artificial Intelligence (AI) is transformative, shaking up entire industries with powerful new machine learning (ML) algorithms enabled by abundant data, fast hardware and cloud computing. But investment in AI also carries risks, from bad data to poorly validated models to insufficient understanding of how best to use these tools and integrate their outputs into decisions.

This white paper is an initial effort to apprise relevant stakeholders of the capabilities and risks of AI, and of best practices in AI product development and broader organizational "AI readiness". As well, it aims to provide an early glimpse of how AI might be used to create, enhance and automate decision support systems in labour market data applications. (A more detailed solution architecture will follow this work).

Based on initial research on comparable work, and interviews with regional stakeholders and experts, the outlook is promising: The right combination of data, visualization, analytics and AI can likely provide a powerful value proposition for multiple stakeholders in the Eastern Ontario economic development, business/employer, labour market services, post-secondary education, funders and policymakers communities.

If a comprehensive solution addresses not only the technical requirements of an analytics system, but also any broader AI readiness and data literacy gaps in the decision ecosystem, the results should be:

- Increased capacity for decision makers in Eastern Ontario to make data-informed decisions when launching new labour market initiatives, business attraction programs, training curricula, regionally beneficial advocacy campaigns, and other initiatives; and
- Increased utilization, productivity and value-add of analysts, allowing them to focus less on low-level "grunt work" and more on high-level forecasts, implications and recommendations that address questions and issues most often asked by senior decision-makers.

Contents

1. Background and Objectives

The Eastern Ontario Leadership Council is undertaking an ambitious project aimed ultimately at the creation of an AI-powered data analytics and forecasting capability that would assist the region's governments, economic development, business and community leaders in making better-informed decisions about labour market priorities and related initiatives to advance the regional economy.

The project leaders engaged a Toronto-based design consulting firm, The Moment, to facilitate stakeholder discussions and exercises towards the assessment of demand, functional needs and technical requirements for a possible product and/or service to deliver those advanced labour analytics capabilities.

The main purpose of this white paper is to inform project stakeholders — including economic development, labour market and community leaders, the consultants, project managers and ultimately those who write and respond to a product/service RFP — of important issues and critical success factors in developing, procuring and using AI systems. It is intended to *align with* the process and the findings of the design and facilitation consultants, but serves a distinct purpose and stands separate from both their final report and from any subsequent detailed technical product requirements or solution architecture document.

Although many of the principles and issues discussed are general — applying to most AI systems across all application domains and sectors — some special attention is given to the context of data analytics platforms and application to employment, training, economic development and related public policy spheres.

In procuring software systems of any kind, it is important not to be overly specific as to particular algorithms or methods. It is much better practice to work from business requirements and desired user experience to features and functionality and then let the engineers decide on which data structures and algorithms best meet those needs. Why? That is, why shouldn't an executive ask for a "convolutional neural network" (for example)? First, because he/she may not really know what works best. Second, because new methods are invented almost daily, and this year's best method may be superseded by next year's, even as the functional requirement (say, "forecast the demand for skilled electricians in Brockville in 5 years") may remain. Nonetheless, a basic high-level familiarity with the methods of AI and data science, their power and limitations, is useful for all stakeholders.

In procuring software systems of any kind, it is important not to be overly specific as to particular algorithms or methods.

Eastern Ontario's Leadership Council Economic Development

2. AI, ML, Data Science and Analytics

AI Today: Rise of the Machines

AI is big. Computer scientist and Coursera co-founder Andrew NG of believes that *"AI is the new electricity"*[i]. Bill Gates opined that a breakthrough in machine learning *"could be worth ten Microsofts"*[ii].

By acting like a capital-labor hybrid, Artificial Intelligence offers the ability to amplify and transcend the current capacity of capital and labor to propel economic growth. Accenture estimates that greater absorption of AI into all sectors could nearly double economic growth from now through 2035 in several major industrial nations and extend Canada's baseline growth by 60 percent.[iii] Eastern Ontario has articulated a desire to keep pace in this and other technology deployment arenas (e.g., specialized applications of ecommerce, connected autonomous vehicles, start-up / first customer matching platform).

The AI Revolution: Drivers, Enablers, Obstacles

What is driving the astonishing growth in AI investments, development and adoption? To put it simply, ***AI is making prediction cheap***. In their 2018 book[iv], *"Prediction Machines"*, authors Ajay Agrawal, Joshua Gans and Avi Goldfarb compare the AI revolution of our era to the semiconductor revolution of the 1970s-90s.

During those earlier decades, faster, smaller and ever-cheaper computer "chips" reduced the cost of calculation and allowed us to embed calculation into many devices and systems — including cars, home appliances, weapons, phones and iPods. They made everyday objects **"smarter".**

Benefits of Artificial Intelligence for Decision-Making

Faster + Smarter → More Cost-effective

Similarly, the proliferation of AI algorithms and trained machine learning models allows us to infuse "intelligence" — in the form of prediction, recognition and reasoning — into our devices, appliances, systems, vehicles, infrastructure and into nearly every corner of our increasingly digital economy.

This automated prediction, recognition and reasoning is valuable because it reduces uncertainty; and uncertainty is costly in business, engineering, medicine, public policy and other spheres.

Consider some examples. If you're in sales, you typically want to know:

- Which of your customers are most likely to buy an upgrade?
- How much would they pay?
- Which customers are most likely to leave you for a competitor's product? When?

To have even "pretty good" answers to these questions can help an executive to focus marketing and sales budgets where it matters most, and to charge prices that maximize total expected revenues.

Some more examples:

- A security system recognizing faces of known criminals in a bank or airport.
- An "automated receptionist" at a telemedicine centre, understanding the typed or spoken words of a patient, and contacting the appropriate medical professional.
- A Bay Street trader's "intelligent assistant" predicting currency fluctuations.
- A factory information system predicting when a crucial machine component will likely fail.

The possibilities are almost endless for reducing uncertainty to make better business, medical, engineering, or policy decisions. Those organizations that do not adopt digital transformation, data science and AI risk falling behind, due to poor decision-making, missed opportunities and negative perceptions in the market.

If cheap prediction, classification and recognition are driving the revolution, what is enabling it? In other words, why *now*, instead of in 1990 or 2040?

First, there is abundant data for "training" many AI models used today. The rise of social media, cloud data storage, open data initiatives, electronic publishing, and internet-connected devices all contribute to a veritable tsunami of data. Roughly 2.5 quintillion (10^{18}) bytes of data are generated each day — and rising; the world's stored data doubles every two years.[v] (Despite an overall abundance, gaps and inequities in data availability and affordability exist, especially for smaller and economically disadvantaged communities).

Second, computing hardware is faster and cheaper. Graphics processing units (GPUs) that were designed primarily to animate video game displays are great at doing the fast vector and matrix arithmetic required by artificial neural networks and other machine learning algorithms. There are also special "neuromorphic chips" designed patterned after brain structure and intended for use on today's deep neural network algorithms. As the cost of GPUs, computer memory and other hardware has come down, more companies, schools, local governments and citizens can own powerful computers. And with the cloud, one doesn't even need to own it — one essentially *rents* supercomputing capabilities.

Third, the algorithms are better. The ***deep learning*** paradigm — pioneered by Canadian researchers Geoffrey Hinton and Yoshua Bengio, among others — has greatly improved our ability to train AI models that perform at or near human expert levels in everything from chess to facial recognition to language translation to bond trading. A vast ecosystem of inventors and developers continue to improve, teach and learn, assisted by ever better tools for designing, coding and testing. (Are there limits? Almost certainly. No engineered AI system has yet achieved the robust real-world intellectual breadth and agility of a 10-year-old human child. Experts are divided on whether or when they will).

So the proverbial playing field becomes more level, through access to the internet, cloud computing, open-source software, and free or cheap online instruction — it is amazing what good AI/ML materials can be found on GitHub, Google Colab and YouTube! Of course, where there are drivers and enablers of innovation, there are bound to be obstacles as well. Some technical, legal and policy barriers exist; other brakes on innovation are more a matter of business-as-usual culture, risk-averse managers and knowledge gaps. These are addressed in the section on AI Readiness below.

Beyond the Buzzwords: What is AI?

AI is powerful and transformative, but it isn't magic; so let's demystify it a bit here. The following definitions are the author's; they are not canonical or universally accepted, but they seem to roughly align with the understanding of most experts in the field.

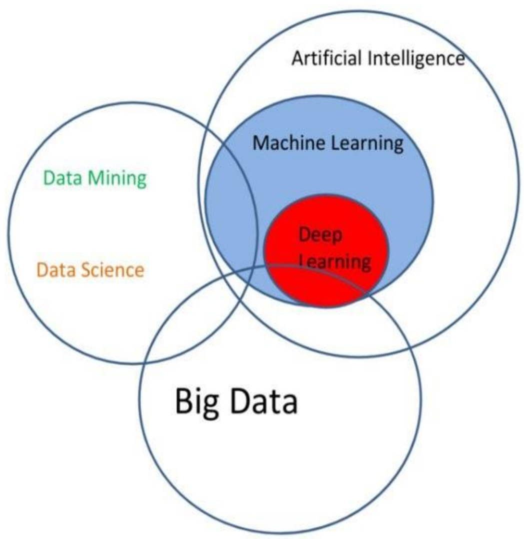

Figure 1: One reasonable view of how various AI- and data-related practices are related.

Data Science *is, broadly, the theory and practice of managing data and extracting meaning and value from it.* ***Data mining*** *is a term from the 1990s that describes methods for extracting meaning and value (like extracting gold from ore).* ***Big data*** *is a term from the past decade that refers to methods for dealing with data at especially large volumes and/or velocity.* Data science includes ***data analytics*** methods, which are typically ***descriptive, predictive or prescriptive***. (The latter two are usually considered part of machine learning).

Artificial Intelligence (A) *is the codification of what we increasingly know about perception, reasoning and learning, and the engineering of those principles into machines.*

Machine learning (ML) *is the science, practice and business of building systems that improve their performance over time.* ***Deep learning*** *is a relatively recent neural network approach that employs many "hidden layers" of "neurons" to learn intermediate representations on the way to producing classifications or predictions.*

The term "***algorithm"*** used throughout this paper refers to a set of logical steps used to solve a problem or perform an action. Software developers implement algorithms by coding them in a programming language such as *C++* or *Python*.

Important task-focused sub-fields of AI include ***computer vision***, ***natural language processing*** (and closely related, ***speech recognition***) and ***robotics***.

It is best not to obsess on the buzzwords *du jour*, but rather to be precise on what problem needs to be solved and then to find the appropriate data, representations and algorithms to best solve the problem.

Again, there isn't universal agreement on these terms, whether and how to use them. The boundaries between sub-disciplines are fuzzy, as they are between AI/ML and the fields of statistics, neuroscience, business analytics and linguistics — all of which have contributed ideas and methods to AI, and *vice versa*. Terminologies and methodologies go in and out of fashion — "pattern recognition" and "expert systems" were popular in their day but are rarely heard today.

The next section provides a very brief summary of important methods; more comprehensive information is available in the cited references. (McKinsey's "Executive's Guide to AI"[vi] , available online, is a good, concise introduction).

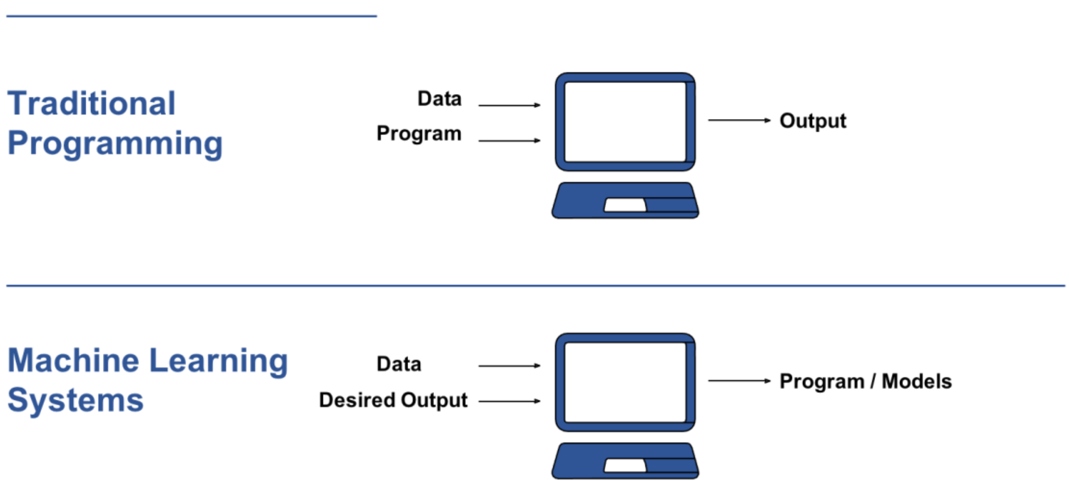

Figure 2: In traditional computer programming, you write a program to take in data and produce particular outputs. In ML, you provide input data and desired outputs to produce a "trained" model that can then be used for classification or prediction when given new inputs..

Some Important Machine Learning Methods

Machine learning tasks are usually divided into *supervised learning, unsupervised learning, and reinforcement learning.*

In ***supervised learning***, training data is divided into inputs and outputs. The goal is to learn a mapping from input to output; that it, to **train a model** that will be able to produce "the right" output for any given vector of inputs. Supervised learning tasks are typically either *regression* or *classification*.

- In ***classification***, the outputs are 1 of N, representing N classes. For example, in speech recognition, the task is to classify a sound pattern as a known word or phrase. In medical diagnosis, the task might be to classify a set of symptoms as a known disease. In face recognition, the task is to identify an image as belonging to a particular person whose picture is stored in a database. If N=2, then it is a *binary classification* task, as in {*healthy, sick*} for a patient, or {*valid, fraudulent*} for a credit card transaction.

Decision trees, support vector machines (SVMs) and *multi-layer neural networks* of various architectures are among the most popular classification methods.

- In ***regression*** models, the outputs are real-valued (i.e., decimal numbers, sometimes normalized to be in the range 0 to 1 or -1 to 1), representing for example predicted bond prices for some future date, or the rating of customers in terms of likelihood to respond to a new marketing campaign.

  Linear and logistic regression methods, and (again) neural networks are commonly used for regression problems.

The “supervised” term above refers to the fact that there is a desired “correct” output pattern for every input pattern shown to the model being trained, as if a human teacher were present in the classroom, supervising while a student drills on spelling or arithmetic examples. For example, if you are training an image recognition system to identify pictures of cats on Facebook, your training data might look like a set of pairs, where each pair contained an input image (perhaps represented as pixels) and an output classification — “1” for “it’s a cat” and “0” meaning “not a cat”. The machine learning algorithm works by modifying the internal parameters of the model so that it gets better and better at providing the right answers, that is, “1” when it is shown a cat and “0” when it is shown something other than a cat.

In ***unsupervised learning***, the goal is not to learn a mapping from inputs to desired outputs; rather, it is to “learn” a new representation of the input data that is advantageous in some way — more compact, lower dimensionality, more informative. Another way to put it is that unsupervised learning methods can help discover the *intrinsic structure* in an otherwise confusing set of data.

Most methods of *clustering* (including *k-means, Gaussian mixtures*) are unsupervised, finding intrinsic structure in the form of “natural” groupings. *Principal component analysis (PCA)* is another unsupervised method, wherein the intrinsic structure of a dataset is described in terms of its directions of highest variance.

A popular use case for unsupervised learning methods is customer segmentation in marketing and sales analysis. Most implementations of *recommender systems* (*“You might like this new movie on Netflix…”*) also use a form of unsupervised machine learning. Finally, such methods are also often used in intermediate “preprocessing” and “feature selection” steps in preparing inputs for training supervised models.

The third major type of machine learning is ***reinforcement learning***, in which a system learns to perform ever better at a task by maximizing “rewards” that it receives for good outcomes. These methods — which helped AI systems beat the best human champions in games like chess and Go — are typically used when there isn’t much training data *per se*, and when it’s hard to define an ideal end state (e.g., the final winning chess board configuration). For this reason reinforcement learning methods are often used in game-playing and in robotics.

Putting It Together: Data Science Workflows

Although recruiters and some professionals take pains to distinguish "data scientists" from "data analysts", "AI developers" and so forth, in practice these fields and their tools and workflows overlap significantly. There is little essential difference between a statistician or data scientist "fitting a non-parametric regression model" and a machine learning engineer "training a neural network**". It is all about using mathematics and computers, logic and statistics, to find and exploit recurring and/or "surprising" or "meaningful" patterns in data.**

This is especially true when AI is applied to the practice of data management and analysis itself, as in the EOLC labour force AI project.

A data analytics platform or environment will typically include tools for managing and visualizing data, analyzing and preprocessing it (including unsupervised ML methods), and for building classifiers and predictive models (using supervised ML methods). Depending on their training and the needs of the moment, a user of such a system may be interacting with high-level summaries and automated analytics modules, or she may be working — sleeves rolled up — deep in the raw data itself.

Data Analytics tools are typically used for:

- Managing & visualizing data
- Analyzing data
- Preprocessing data
- Building classifiers
- Creating predictive models

There is vast variety in workflows, depending on specific goals, assumptions, and the data and tools available, but in general the overall task flow is:
Problem Statement → Data → Models → Decisions.

There are critical success factors and a need for human judgement at many points along this path from raw data to decisions and actions. (Covered in more depth in the AI Readiness section below). There are opportunities to employ AI/ML at multiple points as well. In addition to construction of actual ML models for clustering or classifying data or predicting outcomes, AI can be used to **accelerate and customize data analysis workflows** and to **enhance user experience**.

For example, AI can:

- Automate common workflows
- Provide advice and recommendations *("try a linear regression model on this data"),* based on the user's or other's previous experiences, or on expert knowledge built into the system
- Power chatbots to answer queries and respond to commands in natural language (English, French, etc.)
- Seek out relevant datasets, on a platform, within a private cloud, or on the web
- Parse natural language data (for example, published job listings) into relevant classifications and other data structures
- Assist in data audit/cleaning, by identifying noisy or missing data.

3. AI Readiness and Innovation Capacity: It's Not Just the Algorithms

Developing and Deploying a Machine Learning Model: A Multi-Step Process

There is more to machine learning than the actual machine learning part. For example, typical steps for an AI consulting company in building and delivering a trained ML model include the following:

1. Requirements Analysis
2. Identification of relevant data sources
3. ETL (Extract-Transform-Load) – connection to databases, integrate data for ML
4. Data Audit – Verify quantity and quality needed for ML models
5. Data Preprocessing
6. Model Design and Training — the actual "learning" step!
7. Model Evaluation – Detailed analysis report on accuracy, false positives, false negatives
8. Model Validation – Testing and tuning the model on data not used in training, to better ensure that the trained model "generalizes" well
9. Deployment — turning "research code" code into efficient production code optimized for a particular device and operating context
10. Project Report – Explanation of the code libraries, design rationale, usage and maintenance instructions
11. Knowledge Transfer – Team meetings, walk-through training
12. Post-Deployment Support, Updates.

Also note the need for training, support and updates. As with any software, the task isn't "done" when executable code is produced!

To be aware of this process is to be forewarned about the many places where error, bias and other risks can creep into it. This awareness is useful for anyone involved in procurement, development or use of ML-powered systems.

Skills, Infrastructure and Policies Are Critical Success Factors

Powerful drivers and enablers of digital transformation, data science and AI were described at the top of this white paper. But there are significant obstacles to broader and more successful transformation as well.

Leading consulting firms and management researchers have documented failed digital initiatives and AI projects and identified some of the contributing factors.

As a report[vii] from the MIT-Sloan School of Business put it:

Over the years, we have worked with dozens of companies on their data journeys, advising them on the approaches, techniques, and organizational changes needed to succeed with data, including quality, data science, and AI. From our perspective, these are the two biggest mistakes organizations make:

1. *They underinvest in the organization (people, structure, or culture), process, and the strategic transformations needed to get on offense — in other words, to take full advantage of their data and the data analytics technologies at their disposal.*
2. *They address data quality improperly, which leads them to waste critical resources (time and money) dealing with mundane issues. Bad data, in turn, breeds mistrust in the data, further slowing efforts to create advantage.*

In order to identify and help organizations address such issues, a number of companies, academics and professional organizations have developed survey, audit and assessment tools that go under names like "AI readiness", "data science maturity model" and so forth.

Details vary, but the best of these methodologies look closely not only at hardware and software infrastructure and the obvious hard technical skills, but also at the innovation culture, the incentives, policies and practices that tend to be identified with successful data-driven, AI-powered projects and organizations.

This author (EWS) co-developed an AI Readiness assessment methodology and service[viii] that was used with various public and private organizations in the UAE. The assessment and scoring rationale were roughly as follows and are presented because some of the issues raised are likely to be useful in the Eastern Ontario labour force data context as well.

The core of successful, transformative Artificial Intelligence (AI) is machine learning (ML). The essence of ML is the mastery of three key stages:

- Acquiring, storing, protecting and preparing **data**
- **Analyzing** that data and building **models** that describe, classify or predict
- Deploying those models and **integrating** the models' outputs into business, operations or engineering decisions and actions that achieve an organization's goals.

The methodology therefore assesses an organization and its AI usage according to how well it supports and accomplishes each of those key stages of data science and machine learning practice.

"We are what we repeatedly do. Excellence is not an act, but a habit."

— Aristotle

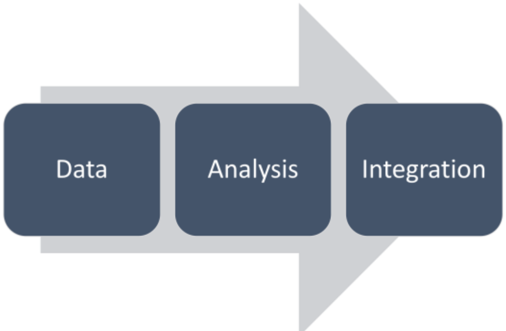

Figure 3: The data science / ML value chain flows from data to analysis and modeling to integration of analysis results into strategic and operational decisions. There are best practices and critical success factors at each stage.

At the same time, an organization's effectiveness in any undertaking is affected by the level and quality of inputs such as **policies and processes** set by management, **available infrastructure** (hardware and software), and the **skills of its people** at different levels and in various departments and roles. This holds true for AI/ML work as well. The methodology therefore also breaks down the assessment of each of the Data, Analysis and Integration phases into questions that focus on these inputs.

AI Readiness – CSFs:

- ✓ Policies and Processes
- ✓ Available Infrastructure
- ✓ Skills

The resulting scoring for an organization can be depicted as a matrix analysis that results in scores across six key areas: *Data, Analysis, Integration; Policy/Process, Infrastructure, Skills.* Within each "box", questionnaires attempt to probe how well the subject organization compares against an idealized "gold standard".

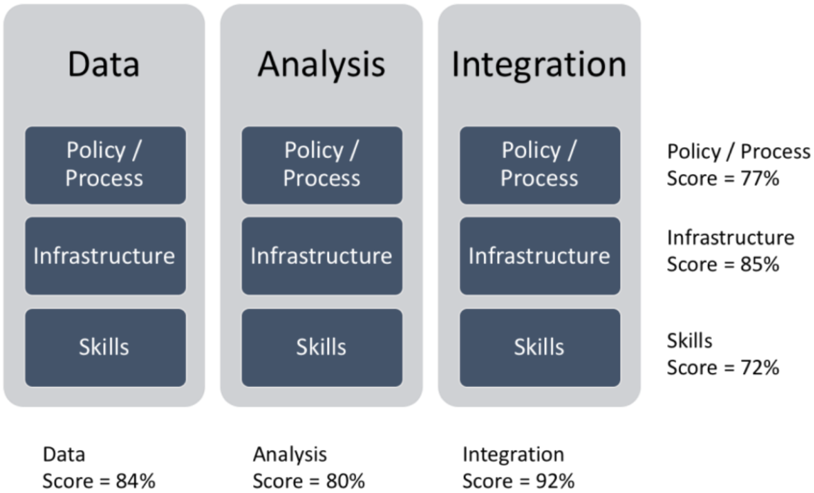

Figure 4: Hypothetical example of an AI Readiness assessment score, based on component scores against gold standards in six critical areas.

Examples of relevant AI Readiness survey questions and issues might include the following (paraphrased):

- *Does management provide financial and non-financial incentives for innovative ideas? Are there hackathons? Is there a data "sandbox"? Is there a clear process around intellectual property protection (invention disclosure, patenting, etc.)?*
- *How many of the executives, managers and other employees have taken courses on AI, data science, or digital transformation?*
- *Is there a formal policy on bias in data and decision-making? On privacy protection?*
- *Describe the computer hardware available on premises (servers, workstations, GPUs). Do you use any cloud services (e.g., AWS, Microsoft Azure)? Describe the physical and digital security mechanisms to prevent unauthorized access to systems and data.*

AI Ethics and Governance

As Cassie Kozyrkov (Chief Decision Scientist at Google) put it, *"When we create machine systems based on data, we teach them a sense of our values."*[ix] One might add, "and our understanding of the world". What data was included in training/building the models? What was left out? Which potential correlative and causal relationships are counted as important? Which ones aren't even considered?

The increasing use of AI in decision-making, and the dependence of those decisions on inherently imperfect data and models has far-reaching implications. The field of "AI Ethics" or "Ethical AI" attempts to reckon with those implications. Within the purview of Ethical AI are considerations of ***Bias, Privacy, Security, Transparency, Displacement and Accountability***, among others.

> *"When we create machine systems based on data, we teach them a sense of our values."*
>
> - C. Kozyrkov, Google

Bias takes several forms in data science / ML work. There are technical concepts such as *sample bias, measurement bias* and *algorithmic bias*, and then there is the reality and perception that a decision system might act somehow unfairly against particular groups of people.

There are algorithmic and statistical ways to detect bias[x], and to correct bias[xi]. To prevent bias in the first place requires clarity, commitment and planning at the earliest stages — during experimental design and data collection. For example, when designing and implementing a survey of jobseekers, one can proactively reach out to indigenous and other historically under-served communities (ex. low-skilled or low-income workers) to ensure that those community members are not under-represented in the survey data and in any subsequent analysis and AI work. Another example of built-in bias in labour-related work: Some jobs are never posted online at all, and therefore wouldn't be found by a web-scraping tool often used to collect data; that absence represents a distortion in any derived understanding of a local or regional labour market.

Privacy and Security are often closely related. Privacy is about policies aimed at protecting potentially confidential or proprietary information; security is about the means to effect that protection. Privacy protection sometimes poses apparent obstacles to data analysis and ML opportunities, or at least requires tradeoffs and clever workarounds, as when synthetic data replaces real medical patient data; when job candidate data becomes unavailable for legal privacy reasons; or when some census variables are left out ("NA") when working down to the neighbourhood or village level of granularity.

Transparency has both a narrow and a broad meaning when it comes to ethical AI. Narrowly, one is concerned with the ***interpretability*** of outputs of a trained ML model or other AI-driven system. In other words, if an AI says, *"Deny this person's loan application"* or *"Launch the missile"*, do we have a way of understanding how the system came to that decision? Or is it a "black box"?

More broadly, stakeholders should demand transparency throughout the entire Data to Analysis to Decision process, including clarity and openness around data provenance (where it comes from, how it was derived) and quality; models chosen (and rejected); model performance and validation results; and how the outputs of a model will be used.

This broader view of transparency brings us to an important conclusion about ethics in data science and AI: Although there are technical approaches to mitigation — for example, statistical methods to detect bias[xii] or to visualize an AI's "decision surface" — the best mitigation strategies start with awareness of the risks and with a commitment to openness, diverse stakeholder inputs, good governance and accountability[xiii].

Sometimes that proactive stance on AI/Data Ethics will be compelled by regulation: For example, the GDPR, the European Union's law on data protection and privacy, has required companies to spend considerable time and resources deciding whether and how to do business in Europe. In the Canadian federal context, 63(3) of the new bill C-11 tabled before Parliament[xiv] has this language that addresses both privacy and transparency:

> **Automated decision system**
>
> **(3)** If the organization has used an automated decision system to make a prediction, recommendation or decision about the individual, the organization must, on request by the individual, provide them with an explanation of the prediction, recommendation or decision and of how the personal information that was used to make the prediction, recommendation or decision was obtained.

Even when not clearly and explicitly required by current law or regulation, industry groups, professional societies and private organizations — ranging from Smart Dubai to the IEEE engineering society to the transnational OECD — are deciding to "get ahead of" any problems by stating their principles and rules of conduct in this area.[xv]

Forewarned is forearmed, and potential AI platform users and decision-makers in Eastern Ontario will have to apply this skeptical awareness not only to their own direct work but also to any datasets and tools that they purchase and use.

4. Concluding Thoughts for the EOLC Labour Force AI Project

Although AI/ML is transforming many industries, it has yet to make major inroads in labour market information (LMI) systems.

The potential value of an AI-driven LMI is vast. Imagine if municipal planners and economic development officers could target company attraction to maximize gainful employment of local graduates and other citizens; could know with high confidence how the Covid-19 pandemic would impact labour supply and demand in their communities. What if college and university leaders could plan new curricula 5 years ahead to more precisely match the expected needs of local and regional employers? What if laid-off workers could reliably identify the retraining that would lead to as-good or even better jobs, without having to move out of their hometown?

Such a system would seamlessly integrate information about jobs openings, required skills, and demand by industries, and about how all of these are affected by economic, technological and demographic trends.

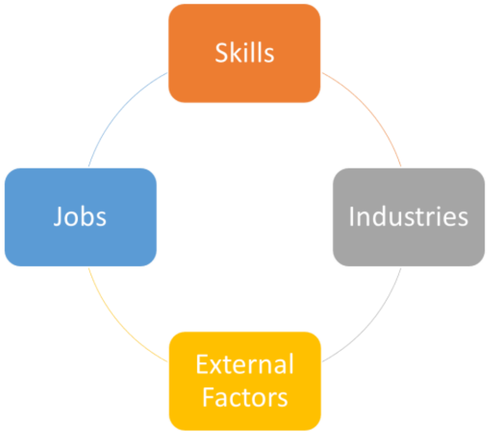

Figure 5: Essential to an EOLC labour force AI solution is some implicit or explicit model of how jobs, skills and industries interrelate, as well as how all these are affected by external social, economic, technological and political factors.

But this is not an easy undertaking. Some of the challenges faced in this application area include[xvi]:

- The high cost of good data
- Geographic granularity — how to find reliable jobs data for small areas (e.g., census sub-divisions)
- Timeliness — how to obtain demographic data (for example) between census years
- Integration of datasets that rely on different definitions, assumptions and formats
- Analysis — paucity of professionals who understand both data science and labour markets
- Data literacy — Even if good data, models, analysis and forecasts are developed, will decision-makers understand and use them to best advantage?

A Phased Approach?

A phased approach may be wise in this situation, to allow time for more of the data literacy and AI readiness factors to be put in place; to grow a community of practice around the needed skills, and the tools and services; and to improve these tools and services through iteration based on user feedback.

A phased approach that seemed to test well in early EOLC stakeholder discussions had the following general structure:

- **Phase 1:** An EOLC "data mart" that provides a level set of relevant, cleaned data along with visualization and analysis tools, a help desk and training for executives and managers in data-driven decision-making. ***Possible Role of AI: Automated parsing of job listings into useful data representations; assists in automating some of the tasks and workflows.***

- **Phase 2:** Add to the data mart customized dashboards featuring current status, recent trends and forecasted future trends in key national, regional and local variables of interest (employment by sector, available workers by occupation, skills in highest demand). ***Possible Role of AI: Powers predictive models; assists in translation between different data "views" for different stakeholder needs.***

- **Phase 3:** Fuller predictive and prescriptive analytics. Decision-makers ask *"What if?"* questions about potential actions and policies around talent attraction, training programs, and so forth, and explore possible outcomes with the assistance of at least semi-automated computational models and scenarios. (Comparable to the Limestone Analytics/EOLC *Economic Impact Explorer* model[xvii] and app, which allows users to view both optimistic and pessimistic economic scenarios in terms of effects on employment, households, and industries.) ***Possible Role of AI: Enables powerful predictive capabilities; supports scenario analysis; permits natural language queries.***

Each version of the product / service must provide at least some real incremental value to all stakeholders (or a lot of value to a few).

Of course, phases can overlap; and other sequences are possible. If cost is an issue, the EOLC purchasers can initially focus the development of more advanced models on just a few variables or questions. The EOLC may wish to develop a solution first for only the region's economic development professionals, for use in targeted campaigns to attract talent to our towns and counties; or for post-secondary institutions, to facilitate curriculum development to meet projected skills needs. The different functionalities could also represent layers instead of phases — a scenarios module built on a forecast-rich dashboard model, and both built on a foundational layer of data and data exploration tools, all delivered more or less simultaneously.

It will be important to keep in mind that a minimal viable product (MVP) and agile product development are based on the idea that each version of the LMI system must provide at least some real value for all relevant stakeholders (or a lot of value for a few).[xviii]'

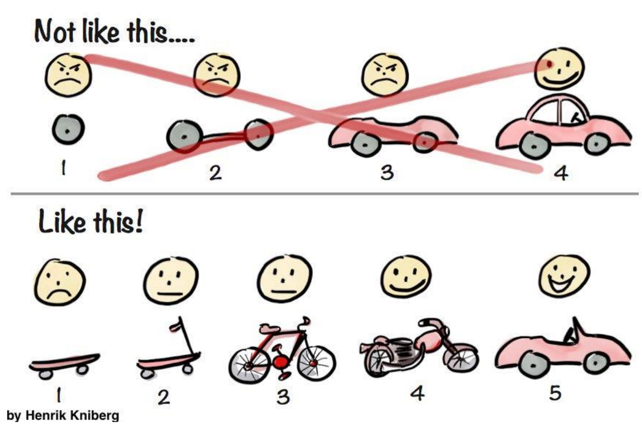

Figure 5: Modern agile software development ensures that at each step along the development path there is a functioning product that provides real value.

A Compelling Example

Elements of the analytics work performed by some local and regional practitioners suggest possible ways forward. For example, Jason Dennison of Fleming College has built an impressive platform[xix] of integrated data, tools and forecasts to help the college decide when to launch new training courses or modify or sunset existing ones in order to best prepare their students and community for future job opportunities.

Among his best practices:

He has a ***coherent and plausible underlying domain model*** that relates future job demand imbalances and external influences (e.g., macro-economy, automation) to job skills for which the College can develop training curricula.

He ***leverages existing tools and data to good advantage*** — Microsoft Teams, Office 360, and Power BI, as well as useful derived datasets such as those from the Brookfield Institute that predict which job categories will be most impacted by automation.[xx]

He ***builds workflows, visualizations and reports that attempt to answer the key questions*** his clients ask when making decisions (in this case about optimizing college course offerings).

Critical Success Factors and Design Constraints

What might all of this teach us about a potentially broader EOLC solution? This report suggests that a broader EOLC solution might be one that would empower economic development officers to attract talent to address the needs of existing or attractable employers); post-secondary institutions to plan courses based on future job needs; and industry leaders to hedge against potential looming shortages of skilled labour? Some initial high-level suggestions follow below. (More details will be available in a subsequent solution architecture document).

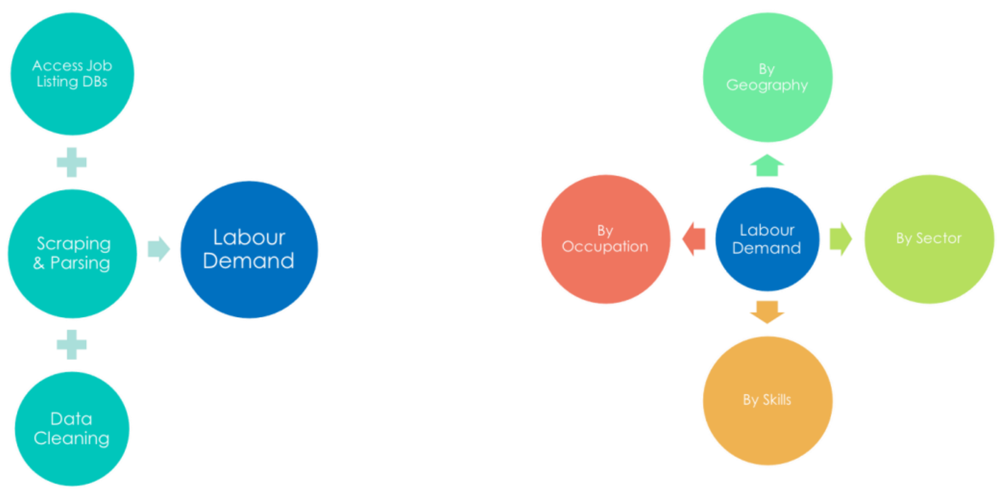

Figure 7: A simple schematic of how a Labour Demand Forecasting module might be set up. Job listings are pulled from multiple online sites and transformed into data structures representing occupations, locales, industry sectors and skill sets — useful to different stakeholders for different planning tasks.

Critical Success Factors & Design Constraints: AI-Powered LMI System

A strong foundational layer of relevant data should be identified and obtained, pooling resources and expertise from among regional municipalities if needed.

The overriding factor in data acquisition should be ***relevance, completeness and quality*** — regardless of whether the providers are public or private, Canadian or foreign-owned.

Resources should be invested in building ***foundational models***. Data models, ontologies, causal models — the technical details can await the next analysis, but the critical point here is that there needs to be an underlying model of how jobs, , occupations, skills and industries inter-relate (as in Figure 7), and how all are affected by exogenous social, technological, economic, environmental and political ("STEEP") factors such as demographic patterns and immigration.

A shared model used by all might be supplemented with more application-specific models and data customized for different types of users — economic development, post-secondary institutions, etc. (See Figure 8).

Interoperability is an important design constraint. The platform should be highly compatible with tools and datasets already used by labour market experts and data analysts in our region (e.g., Power BI, MS Office, EMSI Analyst, O*NET). This can be done through a combination of APIs, data format standardizations and data import/export tools.

Some ***collaboration features are essential***. Users should be able (and encouraged) to save and share favorite datasets, workflows, trained/built models, and analysis results as easily as possible. This will reinforce a growing, learning ***community of practice*** in the region.

The product must be accompanied by support, training and incentives to ensure successful adoption and use of the tools and integration of data-driven, future-aware thinking into decision and planning.

In sum, the right combination of data, visualization, analytics and AI can in theory provide a powerful value proposition for multiple stakeholders in the Eastern Ontario labour, education, economic development and planning communities.

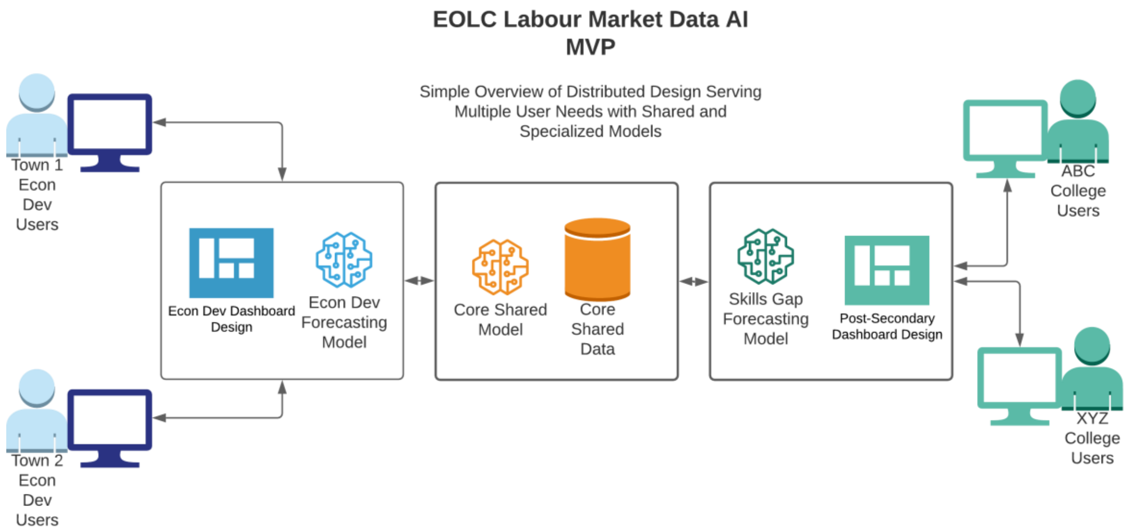

Figure 8: Overview of one possible approach to the EOLC Labour Market Data AI solution, wherein multiple types of users are served by application-specific models built from a common core dataset and skills-jobs-industries model. The data and software would ideally be supplemented with workflow training as well as broader data- and AI-literacy and ethics courses for users and their decision-maker managers and colleagues.

If a comprehensive solution addresses not only the technical requirements of an analytics system, but also any broader "AI readiness" gaps in the decision ecosystem, the results will be:

- Increased capacity for decision makers in Eastern Ontario to make data informed decisions when launching new training programs and business attraction programs; and
- Increased productivity and value-add of analysts, allowing them to focus less on low-level "grunt work" and more on high-level forecasts, implications and recommendations.

Endnotes and References

[i] June 2019, https://www.wipo.int/wipo_magazine/en/2019/03/article_0001.html

[ii] March 2004, New York Times, "Microsoft, Amid Dwindling Interest, Talks Up Computing as a Career"

[iii] September 2016, Accenture report, "Artificial Intelligence is the Future of Growth" https://www.accenture.com/ca-en/insight-artificial-intelligence-future-growth-canada

[iv] 2018, Harvard Business Review Press, ISBN 978-1-63369-567-2

[v] September 2020, EMC and TechJury blog, https://techjury.net/blog/big-data-statistics/#gref

[vi] April 2019, McKinsey & Co., "An Executive's Guide to AI", https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/An%20executives%20guide%20to%20AI/An-executives-guide-to-AI.ashx

[vii] September 2020, MIT-Sloan Management Review, "Getting Serious About Data and Data Science", *https://sloanreview.mit.edu/article/getting-serious-about-data-and-data-science*

[viii] October 2019, Stallion.ai, "AIRA: AI Readiness Assessment", www.stallion.ai

[ix] February 2020, Towards Data Science, "What is `Ground Truth' in AI? (A Warning)", https://towardsdatascience.com/in-ai-the-objective-is-subjective-4614795d179b

[x] May 2018, VentureBeat, "Pymetrics Open Sources Audit AI, an Algorithm Bias Detection Tool", https://venturebeat.com/2018/05/31/pymetrics-open-sources-audit-ai-an-algorithm-bias-detection-tool/

[xi] April 2017, F. Calmon et al., ResearchGate, "Optimized Data Pre-Processing for Discrimination Prevention", https://www.researchgate.net/publication/316030535_Optimized_Data_Pre-Processing_for_Discrimination_Prevention

[xii] May 2019, Brooks Institute report, "Algorithmic Bias Detection and Mitigation: Best Practices and Policies to Reduce Consumer Harms", https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

[xiii] October 2020, Harvard Business Review, "A Practical Guide to Building Ethical AI", https://hbr-org.cdn.ampproject.org/c/s/hbr.org/amp/2020/10/a-practical-guide-to-building-ethical-ai

[xiv] November 2020, House of Commons and Minister of Innovation, Science & Industry, https://parl.ca/DocumentViewer/en/43-2/bill/C-11/first-reading

[xv] February 2020, Thilo Hagendorff, Springer Link. "The Ethics of AI Ethics: An Evaluation of Guidelines", https://link.springer.com/article/10.1007/s11023-020-09517-8

[xvi] September 2018, R. Boselli, M. Cesarini, F. Mercorio, M. Mezzanzanica, Future Generation Computer Systems, “Classifying Online Job Advertisements through Machine Learning”, https://www.sciencedirect.com/science/article/abs/pii/S0167739X17321830

[xvii] Accessed December 2020, Limestone Analytics for EOLC, Covid19 Economic Modelling, http://www.eolc.info/en/working-groups-and-projects/covid-19-economic-modelling.aspx

[xviii] August 2019, Illustration by Henrik Kniberg of Spotify, cited in A. Mersino blog post, https://vitalitychicago.com/blog/my-favorite-agile-illustrations-from-henrik-kniberg-of-spotify-fame/

[xix] October 2020, J. Dennison, Personal Communication (Demo provided over Zoom link).

[xx] June 2017, Brookfield Institute, “Mapping Automation: How will advancing technology impact cities, towns and communities across Canada?”, https://brookfieldinstitute.ca/mapping-automation/