THIS MEMO HAS BEEN PREPARED FOR THE DIGITAL DIRECTORATE.

...

THE ARTIFICIAL INTELLIGENCE UNIT

MEMO

PUBLISHED BY

First printing, 2025

Preface

Data scientists cannot exist outwith the context of an organisation's strategic business outcomes focus, or without the structure that would enable them to investigate, design, develop, deliver, etc., solutions for the outcomes in question.

Hence, this memo is written in the context of overarching areas that may have one or more unambiguous problems that are addressable via machine learning techniques [1]. And, it outlines missing fundamentals [2]; this is especially important because the Scottish Government, and its public bodies, generally underestimate, or ignore, the knowledge, time, and expertise that underpin the automatic & continuous delivery of analytics & machine-learning–dependent solutions via cloud and hybrid infrastructure.

Warning

There is a lot of chatter about productivity and large language models. Especially, inwardly focused solutions that depend on large language models. Yes, an indirect route to improving public-facing services, is by improving internal processes and practices. However, whether a solution improves productivity depends on the value of the work it addresses, and other items. In brief,

$$\text{productivity} = \frac{\textit{value of work}}{\textit{time \& effort expended}} \tag{1}$$

wherein a valuable piece of work is work that enables the achievement of a business outcome goal. The value of a valuable piece of work can be determined via a combination of key business questions and quantitative prioritisation.[1]

This leads to a parallel point. It is absolutely critical, necessary, to define problems first; the article The Most Underrated Skill in Management is an excellent guide.

[1] Machine learning techniques are solution development enablers; if machine learning dictates rather than enables, solutions might have a negative impact on productivity or projects fail [Tabrizi 2019, Panarella 2020, Odilov 2025, Ali 2025].

Finally, there is a fundamental problem in relation to large language models. Organisations are disregarding the fact that the algorithmic architectures of these models - predominantly transformer architectures - are at odds with many of the tasks they are trying to address. In lay terms, it is akin to using a straight line algorithm to represent a cubic relationship; between an input and an output. Over time new approaches will be introduced, meanwhile organisations need to carefully consider investing in, developing solutions based on, the these architectures.

For more details read (a) Emily Bender, e.g., Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data, (b) Gary Marcus, e.g., Taming Silicon Valley: How We Can Ensure That AI Works for Us, Examining LLM's Strengths, Weaknesses, and Future Pathways, (c) GSM-Symbolic: Understanding the Limitations of Mathematical Reasoning in Large Language Models.

Language

The memo adheres to professional/industry definitions of professions and technical terms.

1 Context

Data scientists are neither business intelligence analysts nor statistician, albeit the outputs of business intelligence analysts, of statisticians, are important aspects of data science. Instead, data scientists are problem focused scientists that use a range of artificial intelligence, i.e., machine learning, techniques to develop solutions that **(a)** augment decisions, **(b)** improve productivity, **(c)** anticipate problems, and more.

Data scientists function within the context of an organisation's strategic business outcomes aims. The succeeding sections outline the overarching areas, vis-à-vis the Scottish Government and its public bodies, within which strategic business outcomes are achievable via machine learning dependent solutions.[1]

[1] The term machine learning includes all the deep learning techniques. Almost all artificial intelligence techniques in use today are machine learning techniques.

1.1 Augmenting Decisions, Anticipating Problems, Prevention

Table 1.1 outlines groups of operations, services, and activities that can benefit from decision systems that prevent or anticipate problems, or augment the decisions of professionals. The continuous availability of measures, business intelligence, and statistics is critical to the development of key business questions.

ITEM	HAVE	COMMENT
public/client facing applications	operations, experience, measures	Real-time measures availability aids problem or opportunity identification.
public bodies	activity & event measures	Examples include hospital flow activity measures, telemetry readings of pollutants, bathing water Escherichia coli measures.
robotic process automation lines	operations & experience measures	The focus of robotic process automation is the automation of enterprise services and/or operations, e.g., human resources activities, purchasing & procurement activities, etc. More 1.2.
business intelligence, statistics hub		At present, statistics are more or less hidden.

Table 1.1: Measures, business intelligence, and statistics are critical to deductive and inductive reasoning amongst business teams. It aids the development of key business questions.

Analytics hubs via which business teams can continuously observe measures, metrics, and statistics are critical to deductive and inductive reasoning about services, problems, opportunities, intervention progress, etc. The insights are especially important in relation to inductive, i.e., pattern, reasoning. It is the *patterns* that business professionals and experts progressively or suddenly observe that lead to the **(a)** recognition of a potential problem or opportunity, and **(b)** to the development of Key Business Questions that might benefit from machine learning techniques.[2] Alas, most measures, business intelligence, statistics, etc., are hidden away in Excel, SAS, Access, etc., data files, or in PDF reports. Hindering the organisation's ability to prevent or anticipate problems, or address opportunities, via artificial intelligence, or other, techniques.

[2] For more about Key Business Questions study K Troyanos' articles: 1. Use Data to Answer Your Key Business Questions, 2. How to Make Sure You're Not Using Data Just to Justify Decisions You've Already Made

The prevention of problems is one of the most effective ways to improve services: 1. Public service reform in Scotland: how do we turn rhetoric into reality?, 2. The Preventative State: Rebuilding our local, social and civic foundations, 3. Revenue, capital, prevention: A new public spending framework for the future.

1.2 Machine Learning Augmented Process Automation

Beware, most of the known manual and productivity draining processes of the Scottish Government, including its public bodies, are pure automation problems, i.e., the solutions neither require machine learning nor other augmentations, e.g., operational research. Hence, governments should beware of only funding solutions or teams that address productivity via machine learning dependent solutions.

PHASE	STATUS	COUNT	COMMENT
live	in production	42	Of the 42, 3 are machine learning (ML) augmented solutions, and all 3 ML models are OCR (Optical Character Recognition) models. A range of baseline OCR models are provided by UiPath. The automation team can briefly fine-tune a model if it needs to convert a larger range of text images to text.
live	decommissioned	23	None of the 23 is ML augmented.
live	hypercare	5	1 of the 5 is ML augmented. The ML dependent project is the Redact PII (Personally Identifiable Information) from Freedom of Information (FOI) pilot. As of late last year, the automation team used a combination of (a) an Azure AI Service, and (b) regex, for the project.

Table 1.2: The Azure AI Language Service offers a Personally Identifiable Information Detection Service for redacting a set of categories. An example of regex in practice.

In numbers, consider the set of items listed within the Scottish Government's

UiPath Automation Hub. At present, 466 items are listed. Table 1.2 outlines a few of the items; only a handful are machine learning augmented.

Additionally, a cursory study of a few items also highlights a problem, the lack of an analytics & artificial intelligence core is encouraging teams across the Scottish Government, including its public bodies, to submit analytics request to UiPath (Table 1.3). Some of the the analytics requests might lead to key business questions that might require a machine learning dependent solution.

REQUEST	**TYPE**	**SUBMITTED**	**LIVE DATE**	**PHASE**	**CLIENT**
Immediate Choice - Read & Index Choice Letter (Police & Fire)	RPA + ML[1]	2023-10-26	2025-04-02	LIVE	SPPA
Immediate Choice - Calculate Immediate Choice Options and Generate RSS (Police & Fire)	ANALYTICS	2023-11-03	2025-03-13	LIVE	SPPA
Produce weekly quota reports for the fisher producer organisations	ANALYTICS	2022-12-01		ON HOLD[2]	Marine Analytical Unit, Scottish Government Core
Agricultural Minimum Wage Calculator	ANALYTICS	2022-12-06		AWAITING REVIEW	SASA, ARE, Scottish Government Core
AO (Accountable Officer) Template Process and General approval processes	RPA	2024-03-14		AWAITING REVIEW	ISD, ARE, Scottish Government Core
Request access to SG building	RPA	2024-05-01		SOLUTION DESIGN	Security and Business Continuity, Scottish Government Core
Child Disability Payments - Supporting Information Gathering	RPA	2023-06-26	2024-03-14	LIVE	Social Security Scotland, Executive Agency
Process Certificate of Examination for NSV (National Security Vetting) checks	RPA	2023-02-05		AWAITING REVIEW	Security and Business Continuity, Scottish Government Core
Planning Inspection	EDGE APP, ANALYTICS	2024-07-03		AWAITING REVIEW	Education Scotland, Executive Agency
Risk Management Reporting Automation	ANALYTICS	2022-12-19		AWAITING REVIEW	National Records of Scotland, Non-ministerial offices
Business Continuity Contact List	RPA	2024-03-14		AWAITING REVIEW | ISD, ARE, Scottish Government Core	

Table 1.3: SPPA: Scottish Public Pensions Agency (Executive Agency), RPA: Robotic Process Automation, SASA: Science & Advice for Scottish Agriculture, ARE: Agriculture & Rural Economy, ISD: Information Services Division
[1] Machine Learning Augmented, Optical Character Recognition. [2] The meaning of **on hold** is unclear.

If the focus is

- reducing administrative time costs,
- reducing lead time; the time from the request of a service to its delivery
- simplifying and digitising service request process and steps

in relation to artificial intelligence, then the focus should be the automation team. And because most of the automation problems are pure automation problems, it will be much more effective to contract-in a data scientist as necessary.

If the aim is the direct and indirect improvement of public services via productivity, then some business outcome goals might be achievable via solutions that depend on RPA operational & experience data.

1.3 Large Machine Learning Dependent Systems

These are large scale solutions that should have a staff data scientists. Examples include active metadata systems. They require intensive collaboration with business teams, and others, and a mix of professionals that will be required over time.

1.4 Test Laboratory

A laboratory that

1. Evaluates/tests cyber vulnerability of a model, and of the system it is embedded in.
2. Evaluates/tests the performance of a model from a technical, business, and training datasheet perspective.
3. Evaluates/tests the adversarial vulnerability of a model/system, and more.
4. Evaluates the drift & re-training settings.
5. Serves both public and private sector teams.

The focus will not include the models that the Artificial Intelligence Security Institute focus on.

2 Bare Necessities

Regardless of the strategic focus, the successful delivery of solutions depends on an underlying, and seamlessly integrated network of, infrastructure, technologies, governance and design patterns.

2.1 Starting Off

On a data scientist's or engineer's first day at a private firm, the professional will receive a secure mobile workstation, a secure mobile phone, access to the firm's GitHub Enterprise or GitLab Enterprise account, secure access to the firm's [cloud] computing resources, integrated development environment software details, and much more. In a nutshell, the professional is set up to start understanding a project's landscape, and start contributing, by the second week.[1]

In contrast, vis-à-vis the Scottish Government, acquisition of a mobile workstation can take months. The professionals are responsible for securely setting up their machines; this should not be the case. Individuals individually request permission to purchase a GitHub licence. A development, deployment, and delivery environment does not exist. Et cetera. Imagine the cost of delay in relation to unaddressed problems, and productivity.

Alas, the team of people responsible for these fundamental requirements does not exit. The next section outlines fundamental considerations and requirements.

[1] Project introductions, and meetings with project colleagues, stakeholders, etc., occur during the first week.

2.2 Aims & Structures

Table 2.1 has links to diagrams that outline a sample operating structure that enable analytics and applied artificial intelligence.

item	comment
federated operating model	The federated model illustrates the infrastructure requirements of analytics & artificial intelligence in the context of other business units.
an operating model unit	An example of a unit set-up.

Table 2.1: A model.

The federated model is flexible, avoids business model clashes, aids security, enables innovation, etc. Next, critical functions, and the mix of professionals responsible. The succeeding sections refer to the groups of Table 2.2

Group	Professionals
Cloud	Cloud Engineer, Cloud Architect
Security	Security Architect, Security Engineer, Security Governance Lead
Analytics	Business Intelligence Analyst, Statistician, Data Engineer, Software Engineer, Technical architect, Solution Architect
Analytics Engineering	Data Engineer, Software Engineer
AI/ML	Data Scientist, AI Engineer (AIE)/ML Engineer (MLE), Data Engineer, AI Architect, AI Governance Lead, Data Governance Lead
AI/ML Engineering	Data Scientist, AIE, MLE
Operations	Data Operations (DataOps) Engineer, Development Operations (DevOps) Engineer
Data Architecture	
Governance	AI Governance Lead, Security Governance Lead, Data Governance Lead

Table 2.2: The responsibilities per profession are outlined within the professionals chapter. It is quite probable that the AIE, in collaboration with the DevOps & Cloud Engineers, will address Machine Learning Operations (MlOps).

2.2.1 *Platform*

The key considerations, if the platform is a cloud platform, include

- Platform architecture, which focuses on principles, guardrails, patterns, etc. The key groups are the cloud and security groups.
- Platform engineering, which involves - amongst other responsibilities - setting-up the fundamentals vis-à-vis security and governance, such that functional and business teams can focus on their core design, engineering, delivery responsibilities. The key groups are the cloud and security groups; the operations group does have governance responsibilities also.
- Continuous integrations & delivery: The key groups are the operations, analytics engineering, and AI/ML engineering groups.
- Data architecture: Addressed in collaboration with the organisation's architecture team

2.2.2 *Security*

The responsibilities herein can be overwhelming, Table 2.3, this is why it is critical to have the right people in place.

Task	Description	Comment
Security Governance	1, 2	The cloud professionals must continuously implement best practices, and keep in-line with the organisation's expectations by (a) continuously liaising with the Cyber Security Unit, and (b) keeping abreast of best practices.
Security Assurance		The focus is ascertaining the security and privacy of assets. The key professionals are the cloud, operations, software, data, artificial intelligence engineering professionals.
Threat Detection	1	Cloud & Operations Professionals
Vulnerability Management		The [automatic] continuous & proactive mitigation, remediation, etc., of security vulnerabilities. DEPENDABOT & CodeQL of the security node. AI, Software, Data, and Operations Engineers.
Identity & Access Management (IAM)	1, 2, 3	IAM systems " . . . assign every user a distinct digital identity with permissions that are tailored to the user's role, compliance needs and other factors".
Incident Response	1	All cloud account members must have a basic understanding. Cloud & Operations Professionals.
Data Protection	1	Technology: Amazon Macie
Application Security	1, 2	This is an important aspect of the model & system development life-cycles. Each engineer must be highly cognizant of application security.
Infrastructure Protection	1, 2	

Table 2.3: Security

2.2.3 *Operations*

Responsible for

- Delivering solutions in line with DataOps & Composable Architecture best practices.[**DK2018**]
- Aspects of encoding data and artificial intelligence governance into design, development, deployment, and delivery infrastructures.

Focus	**Reference**	**Comment**
Observability	1, 2	The focus herein is "... maintaining the availability, performance and security of ... software systems and cloud computing environments".
Events, Incident & Problem Management	1	The article Use a process for event, incident, and problem management outlines the import of managing these items.
Operations Design	1	Version control, change & release management, configurations management, build & deployment management, patch management, design standards, etc.
Application Management	1	Application Management is "... *is the practice of overseeing software applications throughout their lifecycle*"
Availability & Continuity	1	

Table 2.4: Operations: DataOps, DevOps, Cloud Engineers, Security Experts

The promulgation of best practices is a key responsibility of operations professionals; collaborating with cloud and security colleagues. For example, the development of application programming interfaces, in-line with best practices, across Scotland's public bodies and Scottish Government Core.

2.2.4 *Product*

Focusing on product engineering & operations. The product teams will either be (a) artificial intelligence teams developing models, (b) analytics teams responsible metrics, business intelligence, and similar, and (c) software engineering teams. For a product in question, it is possible that the project will start-off with an analytics team structure, morph into an artificial intelligence team, and finally morph into a software engineering team.

2.3 Norms: Analytics & Artificial Intelligence

Purposeless data acquisition is not an aim. Data will only be acquired in relation to a problem in focus.

3 Landscape

This brief chapter highlights a few challenges that need to be addressed at senior levels.

3.1 Chief Information Officer

In practice, the Chief Information Officer (CIO) will either write, or intensively oversee, this type of memo. Especially within organisations that are at the starting points of the data, analytics, and artificial intelligence maturity curves. To understand the import of the CIO - usually the head of all other technology, data, and security chiefs - refer to

- Strategy: Chief Information Officer Agenda E-Book
- Mix of Experts (vis-à-vis Cloud Infrastructure): Page 9 of the attached Cloud Migration Roadmap E-Book
- Chief Information Officer: Chief Information Officer Report

which do focus on artificial intelligence.

3.2 Funding

A number of teams have a well-defined, and machine learning (ML) addressable, business problem, lying dormant. The October 2023 Artificial Intelligence Workshop hints at the types of ML (machine learning) addressable problems. However, no team has funding. This problem has not been addressed yet.

3.3 Research/Products

If data scientists are expected to randomly help colleagues trying tools, or exploring data science - the organisation will have to hire dedicated research data scientists for this. Data science is time-consuming, and switching from the extensive and demanding project responsibilities to numerous research enquires will be detrimental to productivity.

3.4 Procured Artificial Intelligence Dependent Software

It is important that leaders officially communicate the overarching responsibilities of teams in relation to procured artificial intelligence dependent software. Otherwise, front-line staff will be continuously, and unnecessarily, bombarded with queries which they cannot address.

4 Professionals & Expertise

Visit Government Digital and Data Profession Capability Framework for definitions, otherwise, a link to a definition is provided - if possible.

4.1 Governance

One of the most important aspects of analytics, and machine learning dependent, projects. Required $\longrightarrow$ professions and training in relation to artificial intelligence governance & ethics, data governance, cybersecurity governance & practice.

4.2 Analytics: Business Intelligence, Statistics

It is critical to understand the very different responsibilities of business intelligence analysts, statisticians, and data scientists; and the knowledge and skills that underpin their professions. For more details study The Field Guide to Data Science. Table 4.1 outlines a possible team composition for an analytics - business intelligence, statistics - project.

ROLE
product manager
solution architect
technical architect
business intelligence analyst
statistician
data engineer
software engineer
Business Systems Analyst

The project team composition will always depend on the project and its deployment goal. The architects and engineers have to collaborate with infrastructure and security professionals.

Table 4.1: An example of a team composition for an analytics - business intelligence, statistics - project, which involves the automatic & continuous delivery of analytics.

4.3 Artificial Intelligence Model Development Team

Table 4.2 links to a (Responsible, Accountable, Consulted, Informed) Matrix of roles critical to machine learning model development. Note, the roles only focus on the model the development part of a machine learning dependent solution. Note, data scientists cannot address/elicit requirements of an overarching system.

PROFESSIONAL GROUP	REFERENCE
Artificial Intelligence Model Development Team	RACI Matrix

Table 4.2: RACI Matrix

4.3.1 *Data Scientists*

An in-depth knowledge of a variety of machine learning subjects, e.g., deep learning, Bayesian machine learning, computer vision, natural language processing, etc., mathematics[1], business acumen, and technical skills. The latter is not beneficial without the first three items.

[1] Especially, calculus, linear algebra, geometry, and probability; calculus and probability aids statistics understandings.

Foremost, if Scottish Government, and its public bodies, are keen to attract & retain data scientists, unambiguous and viable strategies are important. Critically, data scientists should not be recruited without cause, or for addressing random data or technical problems. Especially because data science is a knowledge[2] & skills intensive profession, which requires constant practice and studying, and hence continuous time and financial investment. Expectations include

[2] The playbook's fields of artificial intelligence diagram illustrates this point; time series, or signal, modelling falls within pattern recognition.

- The ability to understand a variety of business problems; per project, this will involve in-depth problem analysis meetings with the business team and/or experts, and further research; and deciding the suitability of machine learning.

- Being one of the lead collaborators vis-à-vis planning, writing a succinct machine learning design document.

- Understanding the importance of, and having the ability to, communicate with business teams [Siegel 2024b, Malone 2020] and engineers.

- The development of a viable model. This is a time-consuming, e.g., in relation to the illustrations, at least two or more architectures were considered per project, and a suite of experiments per architecture.

4.4 Engineering, Operations, Security, Architecture

Underpinning Roles

- Systems Engineer / Projects Manager [Systems Engineering Definition]
- Enterprise Architect & Solution Architect: How Enterprise Architects Close the Gap between Technology and Business
- Data Architect
- Cloud/Infrastructure Engineers
- Security Experts/Architects
- Data Operations Engineers: What is a DataOps Engineer?, DataOps is not just DevOps for data, What is DataOps?
- Development Operations Engineers
- Software Engineers

Cloud/Infrastructure Engineers: Presently, each team with an Amazon Web Services account is either (a) paying consultants to oversee its cloud infrastructure, technical architecture, and security, or (b) team members are acting as engineers and architects. The first is expensive, the second is risky because members are not experts/professionals in these areas.

5 Cost of Delay, Initial Investment

The previous chapter's focus was experts. Leaders might question the cost of delivering through extra experts/professionals, albeit a small number. However, cost should be considered in terms of the cost of delay.[1]

[1] Quantifying Cost of Delay

The government digital service's findings about public sector waste, and the sources of waste, reflects the problems within the Scottish Government. Waste sources include legacy technologies and working practices, *fragmented and underused* data, fragmented technologies, and more. It is a matter of time before the cost of statistics, and other business intelligence metrics, will be subject to intense scrutiny. Especially because the high cost of producing these metrics can be attributed to the aforementioned waste sources.

Here is a rough, bare minimum, estimate of the cost of manually producing *value pipeline* statistics. There are approximately 350 statisticians within the Scottish Government. Assuming each is a $B1$ level practitioner earning $30,558$ per year, and assuming each spends approximately 65% of their time, per annum, producing *value pipeline* statistics. Then a lower boundary cost of Scotland's official & accredited official statistics per annum is approximately

$$350 \times 0.65 \times 30,558 = 6,951,945 \tag{5.1}$$

Note that this does not include costs vis-à-vis pensions, employer national insurance, etc. Additionally, it is quite probable that statisticians spend much more than 65% of their time, per annum, on repetitive tasks.

This approximate £6m figure only applies to 350 statisticians, if we factor in the analysts - there are approximately 250 analysts - then the cost of metrics within the Scottish Government will be quite high. An additional concern, most data practitioners focus on fundamental statistics/metrics/business intelligence, not advanced analytics. The wherewithal to move beyond the fundamentals does not exist.

Hence, delaying, or not acting, is costing millions. Much more than what it would cost to have the appropriate solutions in place.

Considerations

- Upfront investment, cost will be recouped.
- Funding model.
- Change management: There will be a need to address re-training due to changing expertise requirements, new working practices, and more.

Index