AI Readiness Diagnostic Findings

Step 3: Ready Recommendations

Step 3 Overview

To find out how you can benefit from examining your institution through a 'data and AI lens', contact our AI & Data Science team at hello@educateventures.com

• With the rise of the digital workplace, **data** is everywhere. You might have data about your user's or employees **physical environment, their virtual environment, their use of digital resources,** and much more besides

• In addition, the **connections** that exist **between** these things are also a form of data, as are the connections that exist between these things and **the people who are using these tools.**

• Ask yourself: Can I learn about the sort of data I should collect for my challenge from **existing research** on the topic?

• What relevant data can I **currently access?**

• Data can be **unimodal** or **multimodal, quantitative** or **qualitative, structured** or **unstructured**

• It is also important to avoid the **'streetlight effect'** – searching for data under an illuminated spot because that's the only area in which you have **visibility,** rather than searching in the surrounding shade, which happens to be a much larger area

• **Key Takeaway:**

• There is an understandable **reticence** about **data collection:** a worry that people who should not be able to access the data that's been collected will end up **seeing it,** or that it will be **misused.** It is therefore extremely important that **ethics** is at the forefront of our thinking when data collection and collation are being considered

Recommendation: What is data?

SUMMARY: Use some of the existing data in your organisation to double-check you've picked an appropriate challenge with which AI might help

Step 3 in the **AI Readiness Framework** is about **wisely identifying** and **collating data**. We want data to help us with a chosen **challenge** in our organisation. We want to understand the differences between **data types** and **sources,** what we have **access** to, and look at **unimodal** and **multimodal** data. So what is data?

- Data is **everywhere.** It's a collection of **information** and facts, and we encounter it in lots of different ways throughout our lives. There's information in **publications, old fashioned card indexes,** and **online databases.** There's information about **currencies,** there's information through **instrumentation,** there's information from **logs;** data is **everywhere**

- Whether that data is **accessible** to us, or whether it's even of any **interest,** are **different questions**

- How do people use data? For example, in education and training it is used for **grades, attendance, evaluations, heating, lighting, interviews, welfare** and **wellbeing, session lengths, logs...** the list goes on! And data is used by educators, trainers, those working in educational businesses and everyone else across the **educational ecosystem**

- What can data offer? If we think about a possible specific **challenge** that we have chosen in the previous steps of the AI Readiness Framework, we can imagine the example of trying to ensure **the continuity of teaching and learning during the Covid pandemic.**

More accurately, we might want to explore what's happened during the **changes** to teaching and learning as a result, because the usual methods and practice have been **disrupted** in unprecedented ways

- The key question becomes: **what data do I need to explore this?** Lots of practice has been happening **online.** Is it different there? Are there differences between different **sessions?** What's the **quality,** what's the **consistency?**

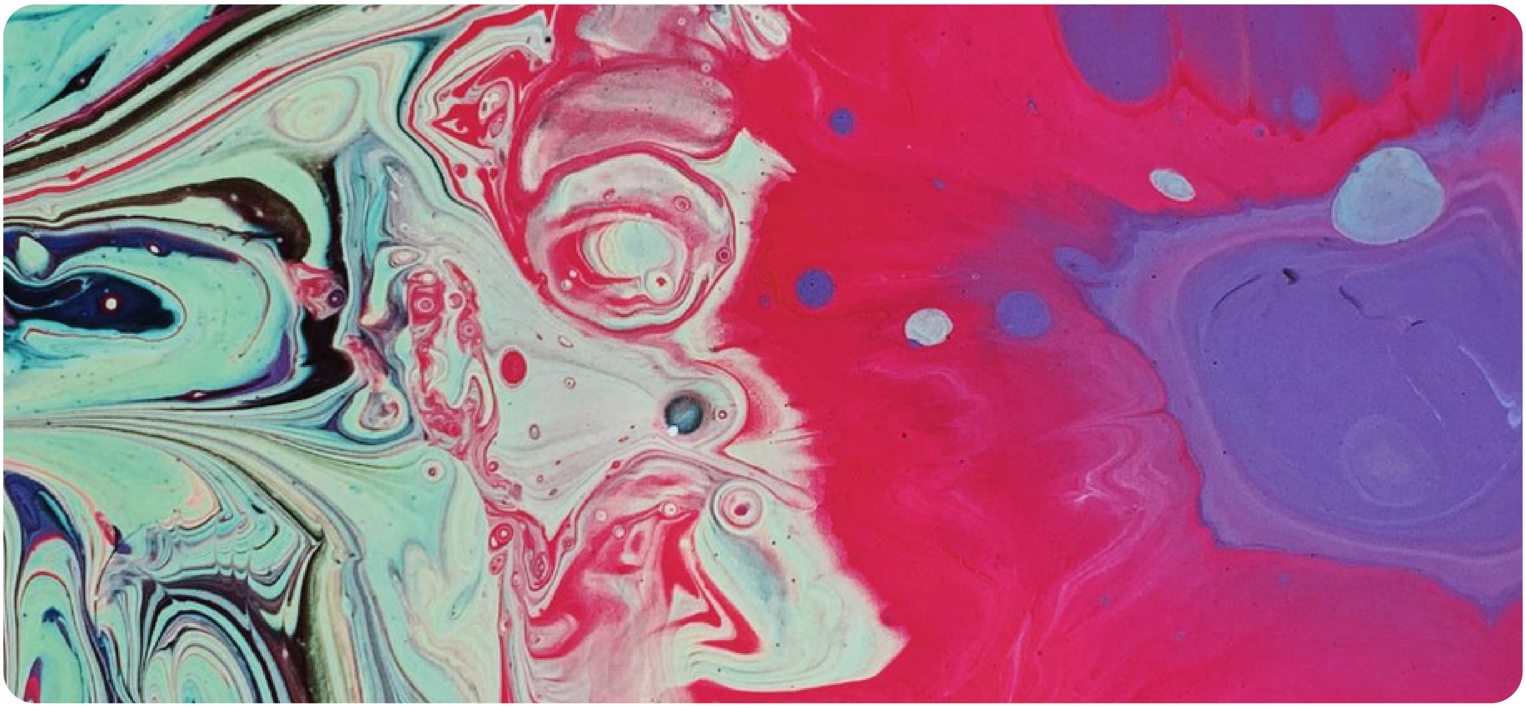
- In order to decide what data we need, **the first thing we should do is:**

- **Look at what data other people have found useful in this situation previously,** because that will give us a clue as to where to look first. There's lots of **research** – not always particularly accessible admittedly – about the ways in which data is often collected, and on the ways data has been useful in the **past**

- **Then we need to think about what's relevant to our challenge,** in order to understand that even if we do have access to data, is it data that we **want** to access in order to address the challenge?

- **Then to ascertain what information between other people suggests what we might have access to,** and if that might be useful for looking at the quality and consistency of teaching and learning in our example

- The **starting point** is all about looking for the data that's going to help us understand our chosen challenge. Looking for data that's already in existence, and thinking about how we can access it. **And even if we want to**



Recommendation: Sources of data

SUMMARY: When considering sources of data, think about where we can get data from, how easily we can get it, who can get it, and its access points

- Where in your organisation might you be able to **access** all this data that you think exists? The starting point is to imagine the **'edges'** of that data, the places that it comes from

- It could be **recorded conversations** from interviews or meetings or Zoom calls, it could be **performance data** from clicks, **answers** to employee wellbeing surveys, training modules, **log data** of interactions with digital interventions, even **eye-tracking data** of eye movements could have been recorded and stored somewhere in your business or school. Eye tracking data sounds pretty advanced but such data capture can monitor how large someone's pupils are, how long their gaze lingers and in what direction. Other data could record how a student is using their hands (are they up from the keyboard or mouse? This is **gesture data**). We could have data about physical aspects of our body, **heart rate, pulse rate** and **temperature,** those kinds of data that can be collected and transferred in real time

- Ask yourself what data you have about people in your organisation. Staff have different levels of **training, contracts.** Where is that information stored? Ask yourself questions about the **physical or digital environment;** we think about a physical environment, is it an office or home? A training site, customer site. Is it travelling? What's going on in the business, do we keep data about **energy consumption?** Or **light and noise levels?** Maybe we have data on the times people **clock in and out** of the building and security and privacy

- Think about **virtual platforms,** think about the kinds of data that they might give us access to about **digital infrastructure and tech stack,** think about resources, the **books,** the **technology,** the **equipment,** the **finance,** all of these places might be points where you'll find that data has been collected

- And then the bit that people often miss when looking very broadly for data is data about the **connections between** the people and the places and the things. Is somebody connected to another, and if so, what's that connection, is it a **collaboration?** How are these different resources connected to each other and how are they connected to the people that addresses. You must think very **broadly** about where you might find data

- Try and look across people, across the places, across the things, and the connections between them, to identify data sources. It's better to identify **more** data sources that are relevant than less, because we can also always **filter out** the ones that aren't **relevant. Try and identify as many data sources as you can**

Recommendation: Accessing and using existing data

SUMMARY: think about what it would take to access your current data, who can access it, and in what form it is and how usable that might be

• It's all very well identifying and discussing the **broad range** of potential data sources that we might be able to access, but what do we know about **accessing** that data, and how **usable** it might be?

• Let's take a **Learning Management System** (LMS) for an example, and examine its data collection and what that really means for people:

• How is that data collected?

• Who owns it?

• Have people consented to the storage of their data on that system?

• Who's responsible for this data?

• How is that responsibility actioned and what checks are in place to ensure safety and accountability is maintained?

• What type of data is it?

• What period do we have that data for? Is it just for a month or a year?

• The data in this hypothetical could be about a particular **course,** and information about it collected from teachers and learners with **consent.** It could be about something that we're interested in that's relevant to our challenge, even

• Is it stored on a central institutional database?

• Is it stored safely, or somewhere else? Hopefully it's stored safely, and hopefully centrally

• Does a course administrator or course tutor have responsibility for this data that people might object to?

• It might involve text or descriptions of subjects' **learning goals, means** and **assessments, teaching methods, resources,** as well as numerical data about **past student performance.** Maybe for several **years.** So we need to know **precisely** what is there, and for what **time period**

• These are the sorts of questions that we get into once we start to explore data and what kind of data is out there. Understanding data access and asking ourselves how accessible it is, is vital. How do you get access to your data, if you need it? Perhaps your colleague who normally processes it is on **leave,** or hasn't followed procedure. Who's responsible for it? Has it been collected **ethically?** Do the people whose data it is **know** that that data has been collected and is being stored about them, possibly to be **used** by you? All of these questions become significant

Recommendation: Introducing multimodal data

SUMMARY: multimodal data is data from multiple sources connected in some way through meaning, providing complimentary information

• Multimodal data is particularly useful when searching for data and is often something people **overlook.** What do we mean by the term **'multimodal'?**

• **Modes** are **channels** of information. We interact with each other and the world through different modes. Multimodal data is therefore data that comes from **multiple sources** that are **connected** in some way through **meaning**, providing **complimentary information** to each other

• This is extremely helpful when looking to analyse data to understand a **problem,** because in trying to address that problem, we want to find different **ways** of looking at it. If we can identify different modes people have used for interacting, for example, then we could collect different sorts of complimentary data about the same **person,** or the same **question,** or **challenge**

• For example:We might have a computer, and we might have somebody who's interacting with it, perhaps capturing some video, and we might also have some some audio connected. Perhaps somebody's singing, for example

• What we really want is to capture all of these **instances of interaction** and to **connect** them to find out something more about our example than just what we might learn through **one** mode of interaction, called **unimodal data**

• If we imagine that our **chosen challenge** (from Step 2 in the AI Readiness Framework) was about determining the quality of the different sorts of interactions occuring in students studying **performance art** online, we might want to be able to gather data about their **movement,** their **speech,** about their **singing,** the way they **interact** with the online technology

• Capturing all of this information would help us to build a picture of not only how **useful** that data is, but how we should go about **addressing** that challenge. We could look for **patterns** in the data to answer questions about how the study of performance has been **conducted online** at the school, college, or university, for example

Recommendation: What multimodal data reveals

SUMMARY: collecting and connecting your data can tell you about what you observe, but it can also tell you about what you can't observe

• Let's think about the sorts of questions that multimodal data can help us to **answer.** We originally asked: what data do I need? And we thought about retrieving examples from other people's **research,** other people's **activity,** that kept by **colleagues.** We then need to think about what our challenge requires, and what we're trying to address, **synthesising** those sets of data in the process

• Consider our data collection is the act of obtaining data about what we can **see:** somebody singing for example. We can **hear** that data, we can watch a **video** of them singing, or of them **walking** into the room, **fidgeting** with the microphone, or where they **look.** We can look at the **log data** about the way a learner interacts with an **intelligent tutoring system,** what they do and don't **click on,** how **often, where** on the page, **what** they're looking at. We can collect data about the **context.** We call this the **input space**

• This is the **observable data** about the way people are behaving in the **context** in which they're behaving

• Some of this data could be captured **automatically,** particularly different sorts of multimodal data. If we consider eye tracking again, that can be captured automatically. We call that **passive capture**

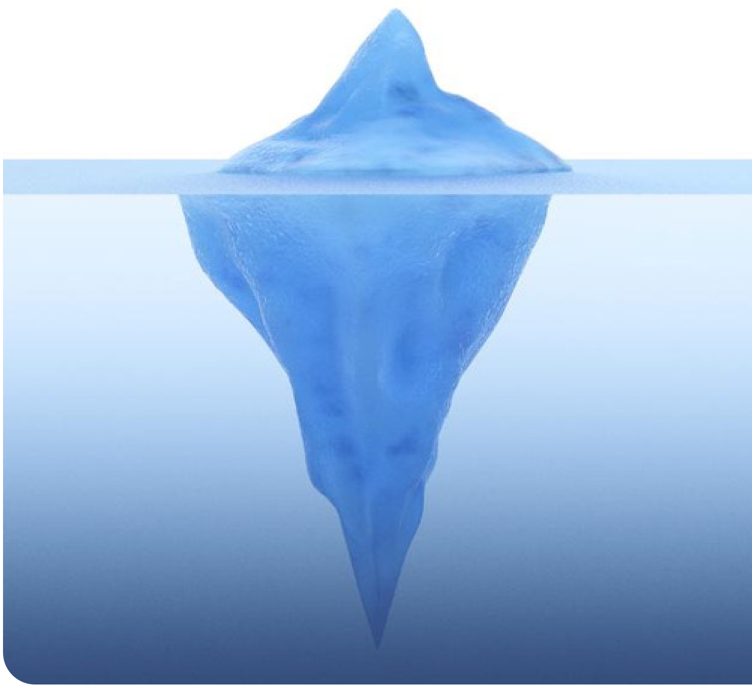
• Lots of this kind of data (heart rate, temperature, breathing) is collected without people consciously, actively **engaging** in the data collection process. It must be done **consenually** of course, but a lot of multimodal data is collected in an automatic way and not as part of **active engagement**

• We can of course gain from this in all the **inferences** and the **interpretations** that we make. Once we've **collected** our data, and **analysed** it, we might start to be able to unpack what's happening with people's **cognition**, their thinking, their learning, their beliefs, their motivation, their emotions

• This becomes the place, if you like, where we can start generating **hypotheses** about what the data might be telling us. We can think about what the data helps us to **infer,** and then move forward with that perspective

• **A significant point emerges here:** yes, we think about what we can observe. That's where we get our data from, but it can tell us about lots of things we **can't observe,** not directly. We can't directly observe somebody's **emotions** changing, but we might be able to collect **contextual data** in which that behaviour occurs, and that will enable us to draw inferences about their emotions. **Well known, well recorded modal data** becomes hugely important

• It is useful, particularly when we're considering using AI on our collected and collated data, to think of that input space as the multimodal data. And our **hypothesis space** where we're going to draw inferences about things we cannot collect data about directly – the space where we add **learning labels.** This is when we start adding information, **annotating** the data to help us to be able to understand things that we can't observe



"**New** data sources can be identified by looking at what is available in our organisation, such as the data sources **related** to ensuring teaching and learning quality. They won't necessarily answer the **immediate** key question we are addressing, but can potentially be used to **complement** and **triangulate,** as well as act as **proxies** to the data that we know we do need to collect."

-- Professor Mutlu Cukurova,
AI Readiness for Educators Step 3

Recommendation: Different data types

SUMMARY: a good place to start is to think about your existing data - is it structured or unstructured? One is a lot easier to understand and search, but the other is more revealing

• One very prominent and commonly found type of data is **structured data**

• This is the kind of data that we see in **height, weight, eye colour, hair.** It's the kind you find in a **database.** It could be s**ocial security numbers, national insurance numbers, postcodes, addresses,** it might be lots and lots of different things depending on who has this structured data. It's what most large companies will probably have about their customers and clients that gives them certain information about **individuals**

• This data is **organised** and relatively easy to **analyse,** but it can **only** tell us so much. It's clearly **defined,** easy to **search,** and straightforward to **analyse** with multiple materials, tools and techniques, that are readily available

• Some data is **unstructured.** It is everything that isn't structured, and actually this is often the sort of information that is particularly **interesting.** This is about **human-generated unstructured data,** such as the **writing** in a letter, a **tweet,** a **document,** the kinds of things we produce for YouTube or in messages you leave to yourself on your mobile. Lots and lots of different data is unstructured, and it's incredibly important. But of course, is is very poorly **defined**

• Unstructured data is **much harder** to search, and **not straightforward** to analyse. It's important we think carefully about the data that we're going to analyse: is it structured or unstructured?

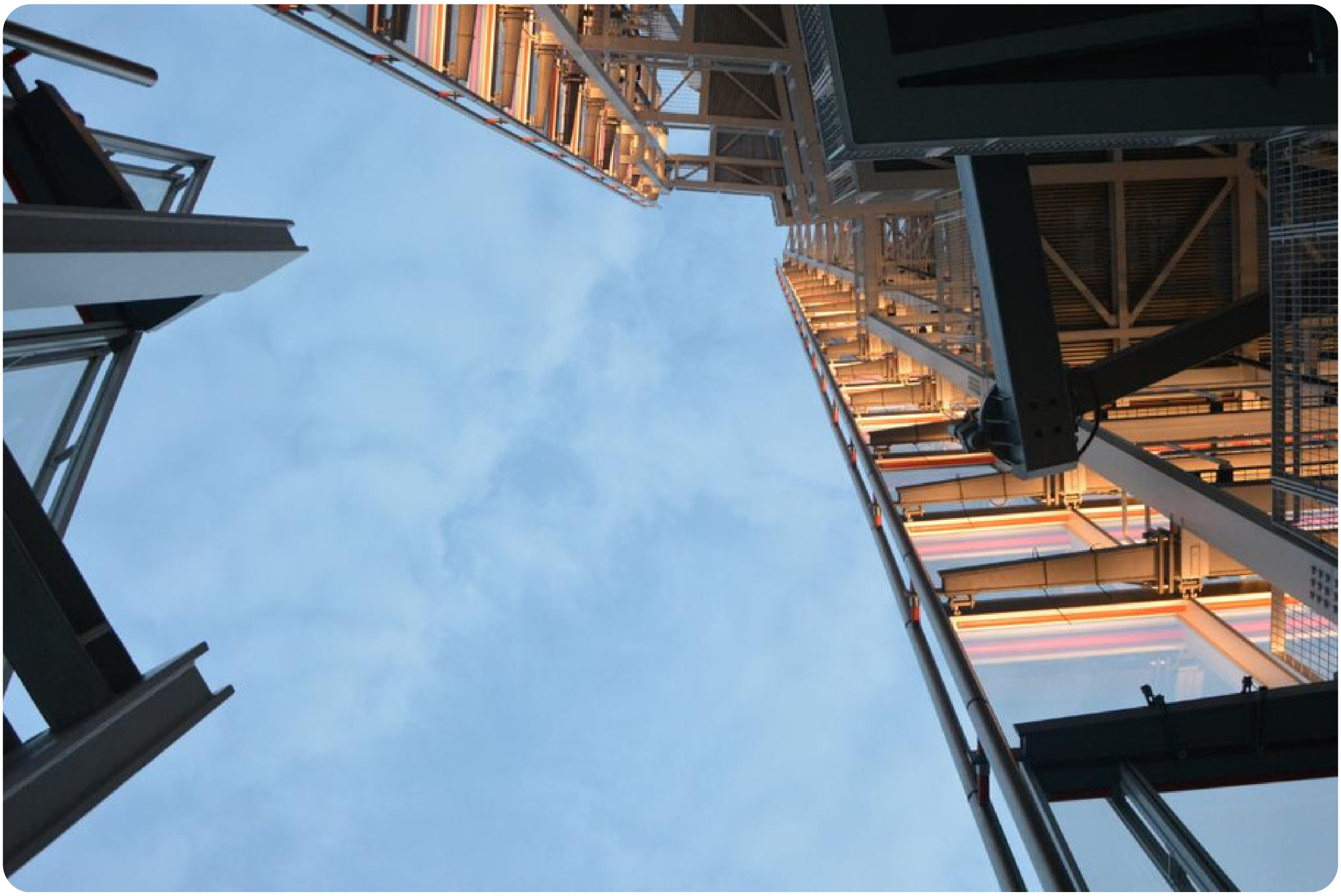
• We can also talk about whether data is **quantitative** (it's about numbers basically) or **qualitative** (this is about how people are feeling, it's about what we're doing, anything that's not quantitative)

• Quantitative data is structured

• Qualitative is unstructured

• Quantitative is much easier to **analyse** than qualitative

• We've asked what kind of different **data sources** we might have access to, we've asked about **channels,** or modes, of information for **multimodal data.** We now ask ourselves about the tyoes of data: quantitative or qualitative? Structured or unstructured? The truth is, we need a **mix**



EDUCATE

Ready to make the most of your data? Find out how at educateventures.com

Recommendation: Multidimensionality

SUMMARY: human intelligence is not static, the relationship with an intervention changes over time or place, for instance, and it is unlikely such changes are simply binary

• The social reality of learning is very, very **complex,** and it is the main subject of most educational organisations. Learning involves complex processes that are a result of a combination of **skills, abilities,** and **knowledge,** and that incorporation of numerous **dimensions** is what we call **multidimensionality**

• If we are going to be able to come up with a **single source of data** able to tell us about something as complex as, for instance, staff wellbeing, we need to figure out what contributes to staff wellbeing and how consistent it is. We might need to work out whether we've got **consistency** in the way our interactions are happening with online teaching and learning, for example. We therefore need to think of our challenges, particularly in the education space, as being **multidimensional**

• Imagine we have some kind of educational interaction, but we want to know what people are **thinking.** How do we find that out? How do we find out if they've understood how the subject matter **fits together?** Are they able to solve the problems? Can they really achieve their best? Or are they really stuck? We need very rich data to help us to understand all these multiple, **connected** challenges

• It's true also that what happens in education is not **static,** it changes over time, and it's highly unlikely that those changes are going to be **binary,** all or nothing. It's **temporal,** in other words; so it's multidimensional and it's temporal, and things happen at different **times,** different **speeds,** in different **order**

• This needs to be taken into account in the **analysis**. The challenges, and the solutions for these challenges, are often very **context-specific.** It is something that gets glossed over when we look at **evidence.** Unless we really understand the **context** in which the evidence was collected, we're not going to know **exactly** how a particular intervention might play out in a different **setting** or with a different **learner**

• Think about a school environment. Think about the enormous differences between a virtual environment, and a physical environment: different contexts for learning. We wouldn't expect the same thing to be happening to the individuals who are taking part in that learning when they might be working **online** as opposed to a traditional situation: **face-to-face.** The process, the outcomes, could be **worse,** and we can't expect the same in different contexts, so: **multidimensional, temporal, contextual**

• **Multimodal data** really helps us to understand the complexity of something like education, and it also helps us to move away from the inevitable enticement of the **'streetlight effect'**

Recommendation: The 'Streetlight Effect'

SUMMARY: a type of observational/cognitive bias, where there's a propensity for people to look for whatever they're after in the most convenient places

• The '**Streetlight Effect**' is a well-known **phenomenon,** typified by a **drunk person** searching for their keys under the glow of a streetlamp at **night.** They are asked why they are searching for their keys in that one small spot when there is so much more **ground** to cover around them, and reply that it is because this is the only place in which they can **see** anything

• The light may be in **one** location, but that does not mean what we are looking for lives there **too**

• We need to use multimodal data in order to help us understand the rich, complex nature of both the **views** and the **context** that exist in education

• We also need to be careful to appreciate a **balancing act:** it's certainly the case that our **unimodal data source** will answer **some** of our questions, but it won't tell us a great deal about the richness of the behaviours and the context outside of the **single data source**

• Multimodal contextual data is complex, and there's a lot of it to deal with. Because it is complex, it is not **transparent.** We could be transparent about individual data sources, but we cannot be transparent about precisely how they all may **interact** with one another, so it's incredibly important to look at the complexities around performance, and to look to model things, which can also be **intrusive**

• Intrusion comes into play if we start to look into people's **contexts,** we're starting to monitor physical phenomena like **heart rate or pulse or temperature,** and it begins to become intrusive. A **survey** for instance, would usually be much less intrusive

• We want to get the **best** data that we possibly can in order to understand our challenge so that we can think about how we might solve it. We want to **balance** the different sorts of data, and the advantages and disadvantages between the kinds of **contextual complexity** we can pick up using multiple data sources, but also the difficulties that we may have in **explaining** exactly what **role** the different data sources are going to have, and precisely how they are going to be used

Who can help me?

We are specialists in ethical AI solutions for schools and education and training businesses - contact our team for help

The EDUCATE AI and Data Science team was formed to consult on and co-design ethical AI solutions to complex problems in data-driven technology ventures and schools. Our team of computer scientists, educationalists, and world-renowned experts can take you from zero AI to a comprehensive evidence- led strategy and beyond, with effective, scalable AI-powered teaching and learning solutions.

To find out how you can benefit from examining your institution through a **'data and AI lens'**, and leveraging the transformational power of AI to tackle your challenges, contact the **AI and Data Science Team** at EDUCATE Ventures Research at **hello@educateventures.com**.

Thanks for reading!

- The EDUCATE Ventures Research

Team Summer 2023

Further Reading

Below you can find a selection of resources, books, podcasts, webinars, and research papers appropriate to your stage of AI Readiness. Good luck!

- AI for School Teachers, Byte-Sized Edition
An easy-to-read 10-page byte-sized summary of the book of the same name, written by Professors Rose Luckin, Mutlu Cukurova, and Headteacher Karine George, members of the senior team actively developing and using the AI Readiness Framework from which these recommendations derive

- How Educators can help Future Learners Outwit the Robots
On Machine Learning and EdTech, Professor Rose Luckin's keynote speech at the Cambridge Summit of Education asked whether education is ready for AI, and suggests how educators can help future learners outwit the robots

- Collecting Data in Schools
Collecting data in schools is known to improve teaching and learning — but what sort of information would be needed for solutions involving the use of Artificial Intelligence?

- What Oak National Academy Usage tells us about education during Covid
A SchoolDash blog covering analysis of Oak's usage data in 2021, providing an example of the kinds of insights possible with the available data

- AI Readiness: Step 3 webinar for Educators/Businesses
Two separate webinars introducing Step 3 of the AI Readiness Framework, one targeted toward educationalists, and the other targeted to educational businesses

- AI for School Teachers
The complete book on the AI Readiness Framework, specifically for teachers and headteachers in schools. It will help teachers and heads understand enough about AI to build a strategy for how it can be used in their school. Though it is pitched to teachers and contains familiar examples, the approach should still be used by education and training organisations working with technology