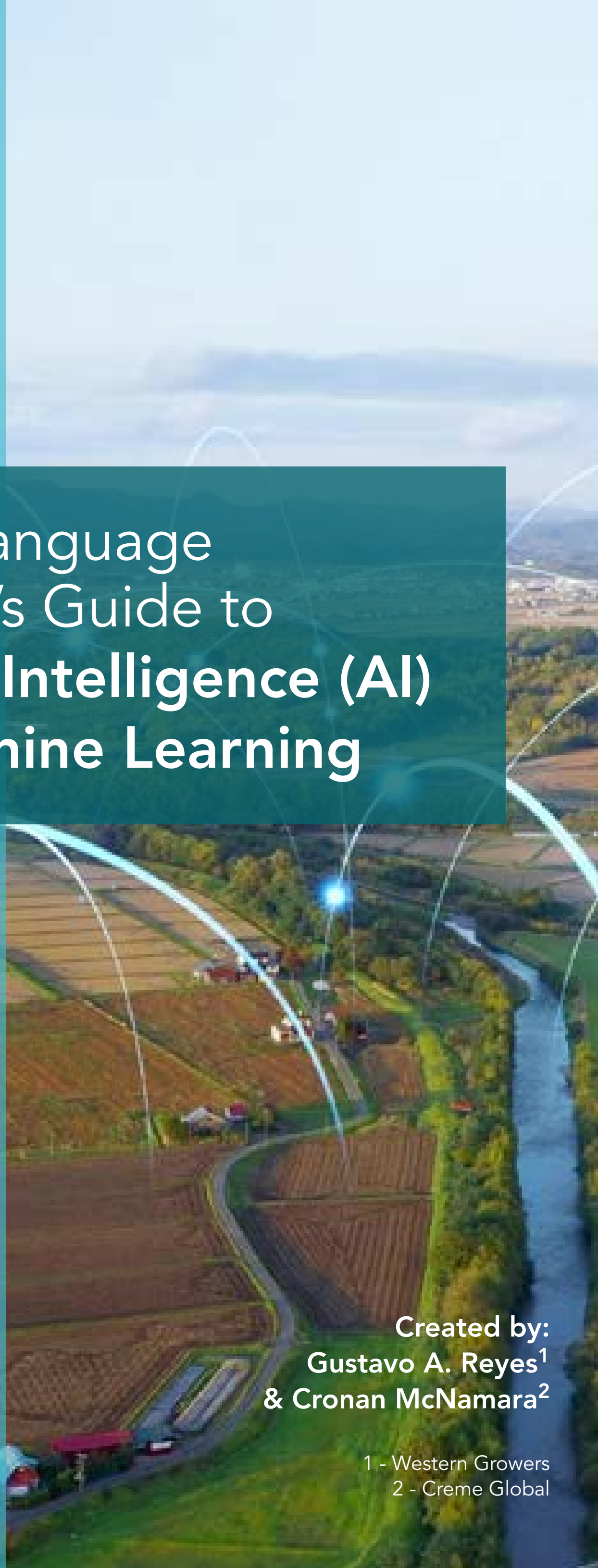


A Plain Language Beginner's Guide to **Artificial Intelligence (AI) and Machine Learning**

WESTERN GROWERS® SCIENCE

Created by:
Gustavo A. Reyes[1]
& Cronan McNamara[2]

1 - Western Growers
2 - Creme Global

TABLE OF CONTENTS

Chapter 1
Introduction

This guide is designed to offer a plain-language, beginner-friendly introduction to the world of Artificial Intelligence (AI) and Machine Learning (ML). Whether you're completely new to these topics or just curious about how they apply to agriculture, food safety and everyday problem-solving, this guide will give you a foundational understanding of the core concepts, common types of AI and how these technologies are already being used in real-world scenarios.

The goal is not only to explain the core concepts of AI, but also to raise awareness about its rapidly growing impact and potential. By the end of this guide, you'll have a foundational understanding that enables you to explore AI further, ask informed questions and think critically about how to adopt and apply it responsibly. When used thoughtfully, AI can significantly enhance how we work, solve problems and drive innovation across everyday operations.

Importantly, this document is not designed to turn you into an AI engineer or machine learning modeler; that would require years of study and many technical books! Instead, the purpose of this guide is to equip you with the right mental model: to understand how AI works at a conceptual level, to recognize where it can be useful in your work and to be an informed participant when your organization adopts or discusses AI tools.

By demystifying AI and showing practical examples, we hope to empower you to:

- Ask the right questions
- Spot opportunities for innovation
- Use AI responsibly
- And feel more confident navigating this fast-evolving space.

Embracing AI should not be scary it should empower us to enhance decision-making, drive efficiency and gather insights.

KEY TAKEAWAYS

AI is about enhancing—not replacing—human decision-making.

The goal is to build a mental model of AI, not become an engineer.

AI is already shaping industries, including agriculture and food safety.

Use AI to ask better questions, spot opportunities and adopt responsibly.

Quick Reference

TERM	DEFINITION
Algorithm	A step-by-step procedure or formula for solving a problem. In ML, it refers to mathematical rules/models used to train on data.
Artificial Intelligence (AI)	The simulation of human intelligence in machines that are programmed to think, reason and solve problems.
Artificial Neural Networks (ANNs)	Computing systems inspired by biological neural networks that can learn complex patterns from data.
Classification	A supervised learning task that predicts discrete categories or labels (e.g., spam vs. not spam).
Deep Learning	A subset of ML that uses artificial neural networks to learn from data in complex ways.
Generative AI (Gen AI)	A type of AI that creates new data, content or simulations based on learned patterns (e.g., ChatGPT).
Machine Learning (ML)	A subset of AI that enables systems to learn from data and improve performance without being explicitly programmed.
Predictive AI outputs	Uses past observations and data to generate forecasts about future events or outcomes (e.g., predicting crop yield or likelihood of contamination).

TERM	DEFINITION
Predictive AI outputs	Builds on predictions by recommending specific actions or strategies to achieve the best outcomes or reduce risks (e.g., suggesting targeted interventions based on predicted contamination risks).
Predictors (Features)	Input variables used by a model to make predictions.
Regression	A supervised learning task that predicts continuous numeric values (e.g., yield, price).
Reinforcement Learning	A learning approach where an agent learns to make decisions by receiving rewards or penalties from its environment.
Responses (Targets)	The outcomes or values that the model is trying to predict.
Root Mean Squared Error (RMSE)	A common metric for evaluating regression models by measuring the average error.
Sensitivity	The true positive rate in classification; how well a model identifies positives.
Specificity	The true negative rate in classification; how well a model identifies negatives.
Supervised Learning	A type of ML where the model is trained on a labeled dataset (with known outputs).
Unsupervised Learning	A type of ML where the model tries to identify patterns in unlabeled data.

Chapter 2
AI Fundamentals

Western Growers Science

What is AI?

While there is no universally agreed-upon definition of Artificial Intelligence (AI), it can be defined as using machines to perform tasks that would require human intelligence.

AI leverages math and statistics to build different types of models that are used to make predictions, recognize patterns, and now, learn and communicate in a human-like form.

Is AI new?

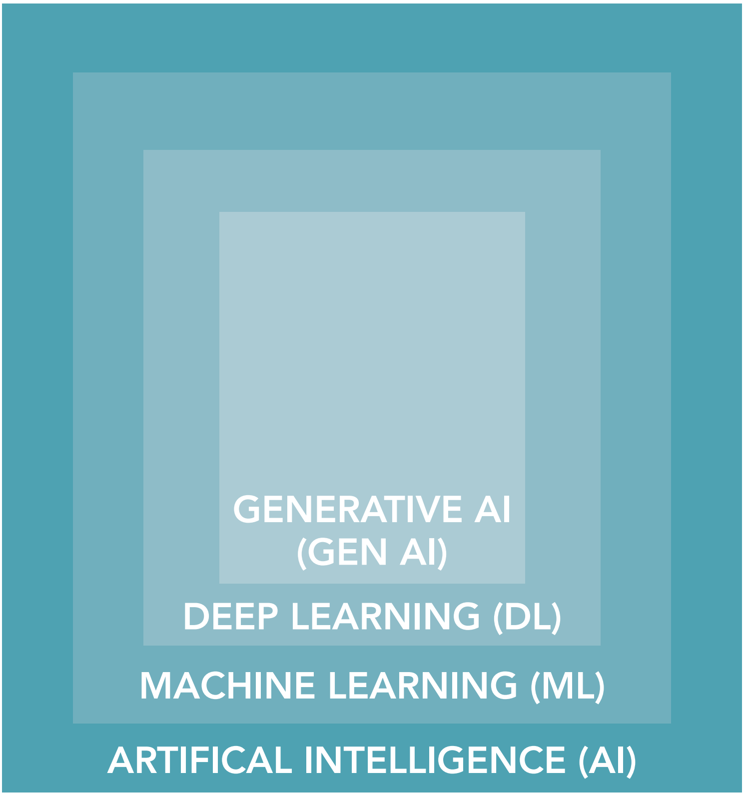
While AI is making headlines more than ever today, it's essential to recognize that AI is not a new phenomenon. For decades, AI, particularly in the form of predictive models, has quietly powered everyday tools and services. From forecasting the weather and estimating crop yields to detecting fraudulent credit card activity, these systems have long worked behind the scenes to make our lives more efficient and informed.

Nevertheless, AI has evolved, and now we often hear the term AI for a more novel use of AI (Generative AI, like ChatGPT). This new type of AI has revolutionized the way we see and use AI. We will dive into this later in this chapter.

The AI Hierarchy

AI is normally categorized by its capabilities and complexity. The field of AI has many layers, with each layer getting more complex and having different characteristics. Here are layers of AI in simple terms:

- **Artificial Intelligence:** As mentioned previously, technology that leverages computers to mimic human behavior (making predictions, recognizing patterns)
- **Machine Learning:** A Subset of AI that takes data and, based on **defined** algorithms, learns from the data and makes to make **predictions**[1] .
    - **Example:** Making predictions of yield based on rainfall.
- **Deep Learning:** Deep Learning is a subset of ML that uses artificial neural networks. These networks consist of many layers that process complex data to identify patterns and make **predictions**[1].
- **Generative AI:** Generative AI is a subset of deep learning that can generate new content, such as text, images or code, based on learned patterns from training data. Tools like ChatGPT and DALL·E are examples of this capability[2.]


GENERATIVE AI
(GEN AI)
DEEP LEARNING (DL)
MACHINE LEARNING (ML)
ARTIFICAL INTELLIGENCE (AI)

An Overview of the AI Layers[1,2,3]

ASPECT	MACHINE LEARNING (ML)	DEEP LEARNING (DL)	GENERATIVE AI (GEN AI)
Definition	Algorithms that learn patterns from data to make predictions or decisions	A subset of ML using neural networks to learn complex patterns	A subset of DL focused on generating new content (text, images, audio, etc.)
Strengths	Simpler models are easier to interpret	Handles large, unstructured data (images, text)	Can create human-like content
	Works well with smaller datasets	High accuracy in complex tasks	Automates creative tasks
Weaknesses	Needs manual feature engineering	Requires large datasets and computer power	Can hallucinate or generate incorrect content
	Limited with complex data types	Harder to interpret	Risk of bias and misuse
Data Needs	Low to moderate	High	Very High

ASPECT	MACHINE LEARNING (ML)	DEEP LEARNING (DL)	GENERATIVE AI (GEN AI)
Compute Needs	Low to moderate	High	Very High
Interpretability	High (especially linear models, decision trees)	Low (neural nets = black box)	Very low (difficult to trace decisions in generative models)
Use Cases	Crop yield prediction	Image Classification	Text Summarization
	Weather Prediction	Speech Recognition	Chatbots and code creation (ChatGPT)
	Fraud detection	Language translation	
Examples	Classification and regression models	TensorFlow, PyTorch	GPT, DALL·E, Copilot

AI-Based Outputs

AI systems can produce different kinds of outputs depending on where in the AI hierarchy we are and what our goal is.

- **Predictive AI Outputs** uses past data to generate forecasts about what is likely to happen next, helping us anticipate risks, outcomes or future trends.
- **Prescriptive AI Outputs** extends predictions by recommending actions based on those predictions, guiding decision-making in real time.
- **Generative AI Outputs**, on the other hand, goes beyond forecasting to create entirely new output, whether text, images, scenarios, that never existed before, offering new possibilities for innovation and problem-solving.

MACHINE LEARNING and DEEP LEARNING		GENERATIVE AI
Predictive AI Outputs "What will happen" (destination time is 24 min)	**Prescriptive AI Outputs** "What will happen" + recommendations based on predictions (destination time is now 25 min, take an alternate route to save 1 min)	**Generative AI Outputs** Text, images, code, interpretations (list of alternate routes, transportation types, price, and cost)

KEY TAKEAWAYS

AI = Machines doing tasks needing human-like intelligence.

ML = Subset of AI where algorithms learn from data.

Deep Learning = Subset of ML using neural networks with many layers.

Generative AI = Creates new content (text, images, code).

AI isn't new—what's new is the **explosion of Generative AI tools.**

Chapter 3

Machine Learning (ML)

WESTERN GROWERS® SCIENCE

As mentioned in Chapter 2, ML is the layer of AI that takes data and, based on defined algorithms*, can make predictions.

Types of Machine Learning

Supervised Learning[4,5,7]

In supervised learning, the model is trained on a labeled dataset, meaning the input data comes with known outcomes. These models can be grouped into different categories depending on their goal. One common distinction is between **regression** and **classification** models:

Regression: used to make predictions on a continuous value (something that falls within a scale of numbers)
Examples:

- Predicting the yield of a crop (30,000 lb. per acre) based on region and rain
- Predicting the value of a home ($540,000) based on the number of bedrooms and neighborhood
- Predicting the MPG of a car (18.5 MPG), based on cylinders, make, model and year.

Classification: used to predict a class or a label, answers if something belongs to a group
Examples:

- Predicting a positive or negative based on region, sample mass and weather
- Sorting if an email is spam or not, based on the email address of the sender, location and time

For this guide, we will be focusing on supervised learning.

*Algorithm: a step-by-step set of instructions designed to solve a specific problem or perform a task. In ML, they are based on math and statistics.

Types of Machine Learning

Unsupervised Learning[6,7]

In unsupervised learning, the model is given **unlabeled data** and must find hidden patterns or structures without any specific outcomes to guide it.

Examples:

- Grouping produce shipments into clusters based on temperature and transit time (clustering)
- Segmenting markets based on consumption patterns
- Dimensionality reduction

Reinforcement Learning

In reinforcement learning, a trial-and-error learning approach is used where an **agent** learns to make decisions by interacting with an environment and receiving rewards or penalties based on its actions.

TYPE	HOW IT WORKS	INPUT DATA	GOAL
Supervised Learning	Learns from labeled data (input + known output)	Labeled (input + outcome)	Predict an output for new data
Unsupervised Learning	Finds patterns or groupings in data without pre-defined labels	Unlabeled (no outcome)	Discover hidden structure in data (clusters)
Reinforcement Learning	Learns by interacting with an environment and receiving rewards/penalties	Feedback from environment	Make a series of decisions to maximize cumulative reward

The Model Structure
(Supervised Learning)[7]

At the heart of the supervised machine learning models, we have a simple idea: use known information (predictors) to make informed guesses (responses).

Predictors (features or inputs): this is the information you already have, the known information you can feed into the model.

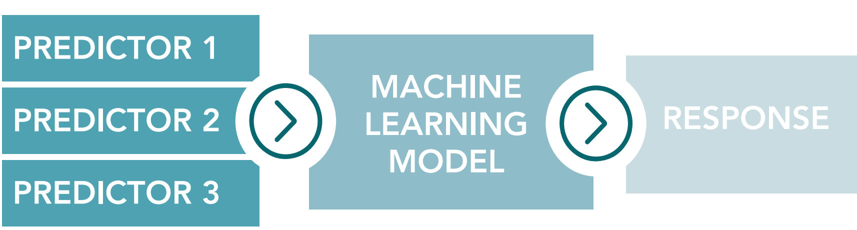
Examples: temperature, soil type, growing information, sample mass

Responses (target or outputs): the value that the model tries to predict based on the predictors; this value is unknown.

Examples: crop yield, positive or negative

Flow of data from predictors to a trained model that generates a prediction (response)

Model is trained on historical labeled data to learn the relationship between predictors and response.



Developing the Model

To develop a supervised learning model, you need a dataset to train your model. This dataset must include both **input features** (the variables or factors used to make predictions) and **labeled outcomes** (the correct answers the model should learn to predict).

The process generally involves the following steps:

FEATURE EXTRACTION
Defining which variable are important

SPLITTING THE DATA
Part goes to training the model, and part goes to testing the model

MODEL TRAINING
Use the training set to fit the model

MODEL EVALUATION
Test the model on the test set and evaluate its accuracy using appropriate metrics

TUNING AND IMPROVEMENT
Adjust parameters or try different algorithms to improve performance

Evaluating the Models

Regression models[7]

Regression models are evaluated based on how closely the predicted values align with the actual values in the test dataset.

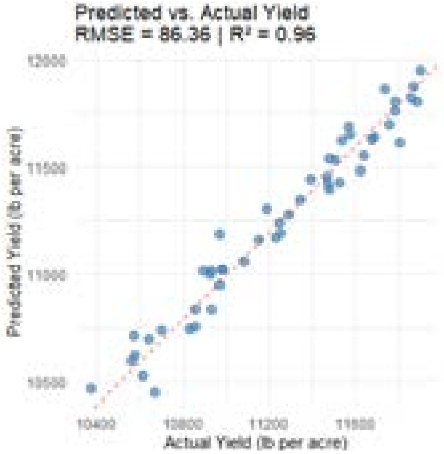
A key performance metric:

- This measures the average magnitude of the errors between predicted and actual values. The lower the RMSE, the more accurate the model.

Other useful metrics include:

- **Mean Absolute Error (MAE):** Gives the average of all absolute errors.
- **R^2 Score:** Indicates how much of the variation in the output is explained by the model.

The plot below visualizes model performance by comparing predicted and actual crop yields, providing a clear view of prediction accuracy through RMSE and R^2 metrics.



This plot shows how closely the predicted crop yields match the actual yields—values near the red dashed line indicate accurate predictions, with a **Root Mean Squared Error (RMSE) of 86.36** and a **coefficient of determination (R^2) of 0.96**, reflecting high model accuracy.

Evaluating the Models

Classification models[7]

Two main measures evaluate classification models

- **Sensitivity** (also called Recall or True Positive Rate): Measures the proportion of actual positives correctly identified by the model.
- **Specificity** (True Negative Rate): Measures the proportion of actual negatives correctly identified.
- In classification models, **both sensitivity and specificity are important**, but the priority depends on the **context of the problem**.
- We want high sensitivity when detecting contaminated food or pathogens, as missing a contamination could lead to a pathogen entering commerce.
- We want high specificity when diagnosing a health condition, as giving a false negative could result in delayed treatment of a condition.

For classification, contingency tables are the best way to visualize these metrics.

		ACTUAL VALUES	
Predicted Values		**Negative**	**Positive**
	Negative	True Negative (Guides Specificity)	False Negative
	Positive	False Positive	True Positive (Guides Sensitivity)

Accuracy: (TP + TN) / Total: How reliable is the model as a whole
Precision: TP / (TP + FP) How reliable a positive prediction is.
Sensitivity: TP/(TP+FN)
Specificity: TN/ (TN+FP)

An example of regression

Let's walk through a practical example to understand how regression works in machine learning. Suppose we want to predict the yield (in pounds per acre) based on different factors. This is a regression problem because the output is a continuous numerical value.

Question: How can we estimate the yield based on weather, soil and regional data?

Training Data (Labeled Data)

FEATURES/PREDICTURES				OUTPUT
Rainfall (inches)	Avg Temp (°F)	Soil pH	Region	Yield (lbs./acre)
12.0	75	6.5	Midwest	28,500
10.5	78	6.8	Southeast	26,200
14.2	72	6.3	Midwest	29,700
9.8	80	7.0	Southwest	24,500
11.5	77	6.7	Southwest	27,300

New Unlabeled Data

FEATURES/PREDICTURES				OUTPUT
Rainfall (inches)	Avg Temp (°F)	Soil pH	Region	Yield (lbs./acre)
14.0	65	6.8	Southwest	????

ML MODEL → Yield Prediction: **28,000**

This diagram illustrates how a supervised machine learning model is trained using historical data, where both the inputs (features) and the outputs (corn yield) are known. Once trained, the model can take new data with similar features (e.g., rainfall, temperature, soil pH and region) and predict the expected yield.

An example for classification

Let's explore a real-world classification example. Suppose we want to **predict whether a crop sample will test positive or negative for contamination** based on environmental and sample characteristics. This is a **classification problem** because the output is a **category or label**, not a continuous number.
Question: Can we classify a crop sample as "Positive" or "Negative" for contamination based on key features?

Training Data (Labeled Data)

FEATURES/PREDICTURES					OUTPUT
Rainfall (inches)	**Avg Temp (°F)**	**Soil pH**	**Region**	**Sample Mass (g)**	**Test Result**
12.0	75	6.5	Midwest	120	Negative
10.5	78	6.8	Southeast	105	Postive
14.2	72	6.3	Midwest	130	Negative
9.8	80	7.0	Southwest	110	Positive

New Unlabeled Data

FEATURES/PREDICTURES					OUTPUT
Rainfall (inches)	**Avg Temp (°F)**	**Soil pH**	**Region**	**Sample Mass (g)**	**Test Result**
12.7	76	6.4	Midwest	118	????

ML MODEL → Result: Negative

The model is trained to recognize patterns between environmental/sample factors and contamination results. Once trained, it can classify new samples as **Positive** or **Negative** based on those inputs.

KEY TAKEAWAYS

Supervised Learning: Learns from labeled data (Regression = numbers, Classification = categories).

Unsupervised Learning: Finds hidden patterns in unlabeled data.

Reinforcement Learning: Learns by trial and error, guided by rewards.

Model evaluation: **Regression → RMSE, R^2; Classification → Accuracy, Sensitivity, Specificity, Precision.**

Key idea: **Feed predictors (inputs) → Model → Response (output prediction).**

Chapter 4

Generative AI

Western Growers Science

Generative AI (Gen AI) is a transformative class of artificial intelligence focused on creating new content. Rather than merely analyzing or predicting based on existing data, generative models generate something novel, whether that’s text, images, audio, videos.

How it works?

Generative AI models are often based on *deep learning* (as we learned in chapter 1), particularly neural networks trained on vast amounts of data. These models learn the patterns, structures and relationships in data to then produce realistic, human-like outputs.

The most well-known generative models today include:

- Large Language Models: Chat-GPT (GPT 4), can generate text and code
- Image generators: DALL.E
- Synthetic data generators: to generate data to simulate predictive models

Potential Use Cases

- **Data analysis:** using this model to generate analysis of data, without having to use specialized software.
- **Report Generation:** Automate the generation of food safety reports, summaries or compliance documents.
- **Predictive Labeling:** Generate likely contamination profiles for unseen conditions or new products, based on data we input to the model.

An example of GenAI use

Supplier	Positive Tests (past 12 months)	Last Audit Score	Water Source	Comments
Field A	2 salmonella test	78%	Surface (untreated)	"Water retest not submitted. Adjacent land near cattle."
Field B	0	95%	Well (treated)	"Consistently compliant. No historical issues."
Field C	1 STEC	88%	Canal (treated)	"Recurring coliforms. Action plan in progress."

AI Prompt

Based on this data, generate a plain-language risk summary and rating (Low, Medium, High) for each field, including reasoning

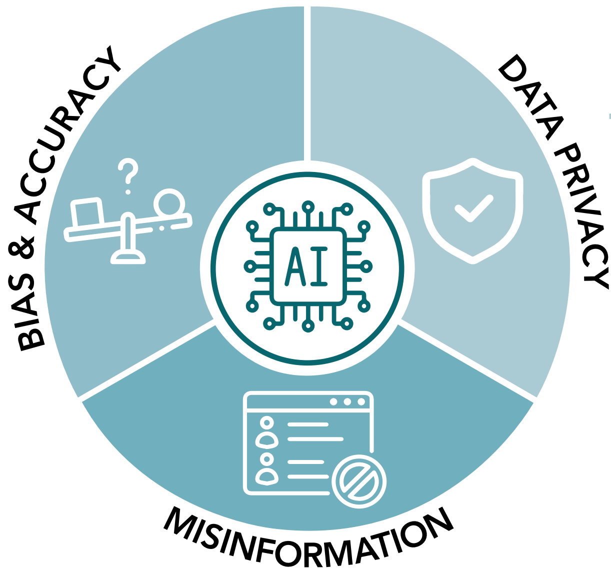
Output

Field A – Risk Level: HIGH Field A shows significant risk due to multiple positive microbial detections (2 salmonella), an audit score below 80 percent, use of untreated surface water, and no water retest was submitted. Proximity to cattle land adds environmental risk.

Responsible use of Gen AI

While powerful, Gen AI comes with ethical and operational considerations

Create an AI policy for your operation



It can generate incorrect or biased outputs if trained on flawed data. E.g. if training data underrepresents a particular region, the outputs may be skewed.

Models trained on sensitive information must follow strict data governance, and a data security evaluation should be conducted.

There's a risk of creating false narratives if not used responsibly.

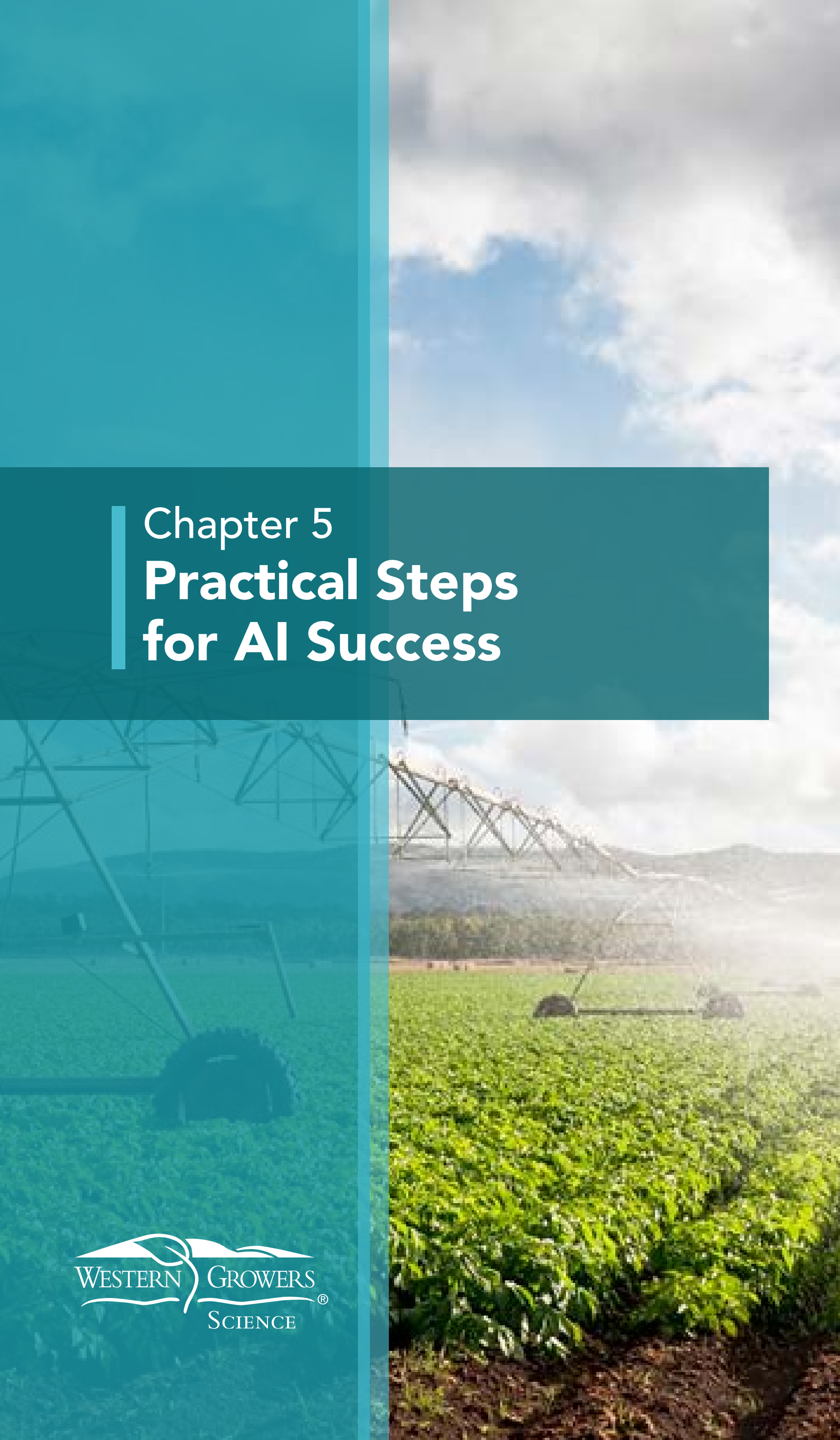
KEY TAKEAWAYS

GenAI goes beyond prediction—
it creates new content.

Examples: ChatGPT (text/code), DALL·E (images),
synthetic data generators.

Use cases: Data analysis, automated reports,
predictive labeling.

Risks: **Bias, privacy issues, misinformation →**
need responsible use and policies.



Chapter 5

Practical Steps for AI Success

You may be thinking, "How do I get started?" In this chapter, we provide a simple AI strategy for you to start utilizing the latest AI tools in your organization, followed by some practical tips to help you use AI effectively and responsibly.

A simple AI strategy to get started in your organization

Step 1. Build your team

- Identify a few potential AI Champions in your team and create a small task force

Step 2. – Secure Private LLM Access

- Have IT sign your team up for Secure Enterprise-level AI model (LLM): e.g. ChatGPT, Gemini, etc.

Step 3 – Create Guardrails

- Establish usage guidelines/governance for the use of AI by the team, simple "do's and don'ts"
- E.g. set up an AI governance team and protocol for your company

Step 4 – Start Experimenting

- Encourage your team(s) to try out these AI tools in their workflow as a collaborator. For example:
- Record meetings
- Write up minutes and actions
- Review or summarize documents
- Write reports

Step 5 – Share and Improve

- Host regular AI Show & Tell sessions where teams share how they're using AI, and celebrate wins.

Tips for Using AI Effectively

- **Mindset –** AI can generate plausible outputs, but not always scientifically accurate ones.
- **Validation –** Triangulate results with other models, team and internet searches.
- **Final Sign-off –** Think of AI as a collaborator and remember it is not infallible. Always have a team member review and sign off on results.
- **Practical Use –** Find repetitive and time-consuming tasks that can be automated.
- **Iteration –** Build a learning loop: Review results frequently and iterate fast.
- **Governance –** Document everything: inputs, model versions, prompts, etc.
- **Culture –** Share results for internal learning.

KEY TAKEAWAYS

Start small and focused: Identify AI Champions and pick 1–2 repetitive tasks to experiment with.

Use trusted, secure tools: Adopt enterprise-level AI platforms and integrate them safely.

Set guardrails early: Establish simple guidelines and governance for responsible use.

Treat AI as a collaborator: Always validate outputs with human review and cross-check sources.

Build a learning loop: Review results regularly, iterate quickly and improve processes.

Document everything: Track inputs, prompts and model versions like you would in software engineering.

Share and celebrate wins: Encourage knowledge sharing to spread adoption and build momentum.

Chapter 6
Conclusion

WESTERN GROWERS® SCIENCE

Artificial Intelligence is not just a technological trend—it is rapidly becoming a foundational tool for solving complex problems, making informed decisions and unlocking efficiencies across nearly every industry, including agriculture and food safety.

Through this guide, you've gained a foundational understanding of:

- AI and Machine Learning
- The types of models used to make predictions or decisions
- The difference between classification and regression
- How AI is already being applied in practical, day-to-day ways
- And the rise of Generative AI and its implications

But this is only the beginning…

What to Do with This Knowledge

- **Be an informed decision-maker.** You now have the language and mental framework to participate in conversations about AI adoption, vendors, tools or policies in your organization.
- **Look for opportunities.** Think about areas where repetitive tasks, large datasets or prediction challenges exist. These are prime opportunities for AI tools to help.
- **Ask the right questions.** Not every problem needs AI, and not all AI tools are created equal. Use your understanding to evaluate where AI adds real value, and where it might introduce risks.
- **Stay curious.** AI is evolving fast. New tools, techniques and ethical challenges emerge regularly. Continue exploring, reading and learning.

KEY TAKEAWAYS

AI is a **foundational tool** for solving complex problems, not just a trend.

You now have the **language and framework** to engage in AI discussions.

Next steps:

1. Be an informed decision-maker.
2. Look for repetitive tasks/datasets AI can help with.
3. Ask the right questions about value vs risk.
4. Stay curious—AI evolves fast.

References/Resources

1. IBM. (n.d.). *AI vs. Machine Learning vs. Deep Learning vs. Neural Networks.* IBM Cloud. https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning
2. IBM Research. (2023). *What is Generative AI?* https://research.ibm.com/blog/what-is-generative-AI
3. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444. https://doi.org/10.1038/nature14539
4. IBM. (2024). *Classification vs. regression: What's the difference?* IBM Cloud Learn Hub. https://www.ibm.com/think/topics/classification-vs-regression
5. Oracle. (2024). *Supervised learning in machine learning*. Oracle. https://www.oracle.com/artificial-intelligence/machine-learning/supervised-learning/
6. Google Cloud. (n.d.). *What is unsupervised learning?* https://cloud.google.com/discover/what-is-unsupervised-learning
7. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An Introduction to Statistical Learning: With Applications in R* (2nd ed.). Springer. https://www.statlearning.com/
8. Regression plot: R Core Team. (2024). *R: A language and environment for statistical computing* (Version 4.3.1) [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Gustavo Reyes, greyes@wga.com