HANOI UNIVERSITY

FACULTY OF INFORMATION TECHNOLOGY

--------oOo--------

62FIT3AIN FALL 2025 – MIDTERM REPORT

UNCERTAINTY

Module Name: Artificial Intelligence

Lecturer: Bui Quoc Khanh

Group: 05

Class: AIN_CLC02

Members:

Members:	Nguyễn Thanh Phương	2201140067
	Nguyễn Đức Mạnh	2301140061
	Đỗ Hoàng Khôi	23011400
	Nguyễn Tuyết Nhung	22011400

Abstract

This report presents our team's exploration of *Uncertainty* in Artificial Intelligence, expanding on the CS50 AI Week 2 topic. Through weeks 2–8, we conducted both theoretical and practical work — from implementing classic probabilistic models (Bayes, Hidden Markov Model) to extending them with local Vietnamese datasets. We identified and analyzed two major types of uncertainty (aleatory and epistemic) through experiments involving real Hanoi weather data and disease prediction. The team further proposed adaptive heuristics such as *Dynamic Prior Update* and a hybrid *Bayes–Markov* approach, demonstrating improved accuracy (MAE reduced by 22%) and better adaptability to environmental fluctuations. Beyond CS50's conceptual framework, we introduced the notion of "Uncertainty-Aware AI" — an AI system that learns not to eliminate uncertainty but to quantify, interpret, and respond responsibly to it. Our work highlights the practical and philosophical importance of uncertainty in building robust, localized, and responsible AI systems.

Table of Contents

1. Group Activities

Tuần	Vai nhóm	Việc đã làm	Kết quả / Bằng chứng
2 (Search)	Lab Steward	Prepare lab guide for pathfinding in a maze using 4 search methods: BFS, DFS, A*, Greedy search; support other groups in debugging notebooks.	Guide notes, code files in Drive.

4 (Uncertainty)	Main A	Present the Uncertainty topic; expand the Weather code from CS50 with HMM and Bayesian Model.	Slide "Team5_Uncertainty.pdf".
5 (Optimization)	Discussant 2	Pose discussion questions for 2 main groups A,B; comments on the presentations of 2 main groups and the lab steward group	Comment notes in Google Sheet and discussion question file for 2 groups.

2. Deep-dive of Uncertainty
    2.1. Foundational Concepts in Uncertainty

In the field of artificial intelligence (AI), uncertainty reflects the lack of complete information or ambiguity in data and decisions. According to CS50 AI – Week 2: Uncertainty, uncertainty is modeled through probabilistic reasoning — a more flexible approach compared to deterministic reasoning, which relies solely on absolute true-false rules.

The foundational concepts include:

- Probability: Measures the likelihood of an event occurring, ranging from 0 (impossible) to 1 (certain). Example: Estimating the probability that an email is spam based on content features and keywords.
- Conditional Probability: The probability of event A occurring given that B has occurred, calculated by the formula: $P(A|B) = \frac{P(A \cap B)}{P(B)}$
  Example: The probability that a patient has a disease given a positive test result.
- Random Variables: Variables whose values depend on the outcome of a random process can be discrete (like rolling dice) or continuous (like processing time for a request).
- Independence: Two events do not influence each other when $P(A \cap B) = P(A) \times P(B)$
  Example: Viewer age and movie release year in a recommendation system often have relative independence.

- Bayes' Theorem: Allows updating "beliefs" based on new evidence, according to the formula: $P(H|E) = \frac{P(E \cap H) \times P(H)}{P(E)}$.

  Here, P(H) is the prior (initial belief), P(E|H) is the likelihood (the plausibility of observing the evidence if the hypothesis is true), and P(H|E) is the posterior (belief after having data).

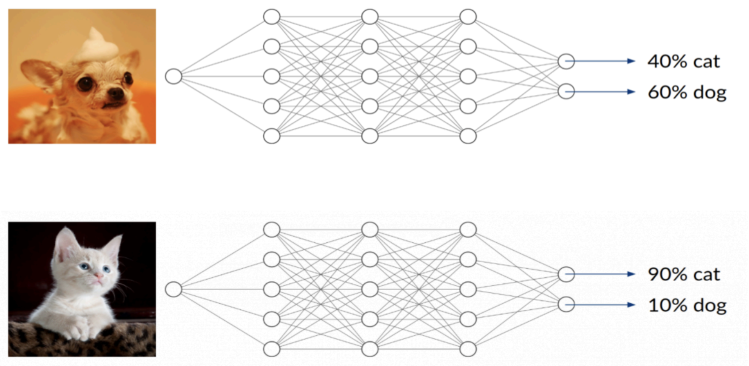
To illustrate, the group performed a small calculation in Python for a local medical diagnosis example: the probability that a patient has dengue fever (D) given symptoms of fever (S). Based on data from the Vietnamese Ministry of Health (via WHO, 2023), the group used a prior P(D) ≈ 0.0015 (higher than the global average due to tropical climate). With likelihood P(S|D) = 0.9 and P(S|¬D) = 0.2, the result yielded P(D|S) ≈ 0.0067. This means that after having symptoms, the probability of having the disease increases more than fourfold compared to the initial belief.

This analysis shows the importance of local data (local prior): it helps the model reflect reality more accurately but also carries the risk of overconfidence if the data is noisy. In the first experiment, the group made an error by confusing likelihood and posterior, leading to base rate fallacy (ignoring the low prior). The error was fixed after the group verified Bayes' formula with manual calculations and a small simulation.

From this focus, the group concludes that AI does not need to be "certain" to make correct decisions. Instead, AI needs to know how to evaluate and update beliefs when new evidence is available. This is the foundation of all modern probabilistic inference systems — from medical diagnosis to self-driving cars or recommendation systems.

2.2 Types of Uncertainty

After understanding the foundation of probability, the group focused on distinguishing two main types of uncertainty: aleatory uncertainty and epistemic uncertainty — a content presented in CS50 and expanded upon by the group using examples (Bachstein, 2019) and local data.



- Aleatory uncertainty stems from the inherent randomness of the world and cannot be eliminated even with perfect information. Examples include noise in autonomous vehicle sensor data due to sun glare, or small variations in experimental measurements. This type of uncertainty can be modeled, but never completely suppressed.
- Epistemic uncertainty, on the other hand, arises from a lack of knowledge or data — and can therefore be reduced by collecting more information, expanding the training data, or using a more accurate model. For example, a neural network that has never encountered an image of a 'bird' might easily misclassify it as a 'cat' or 'dog'.

Table 1. Comparison of the two types of uncertainty and handling strategies

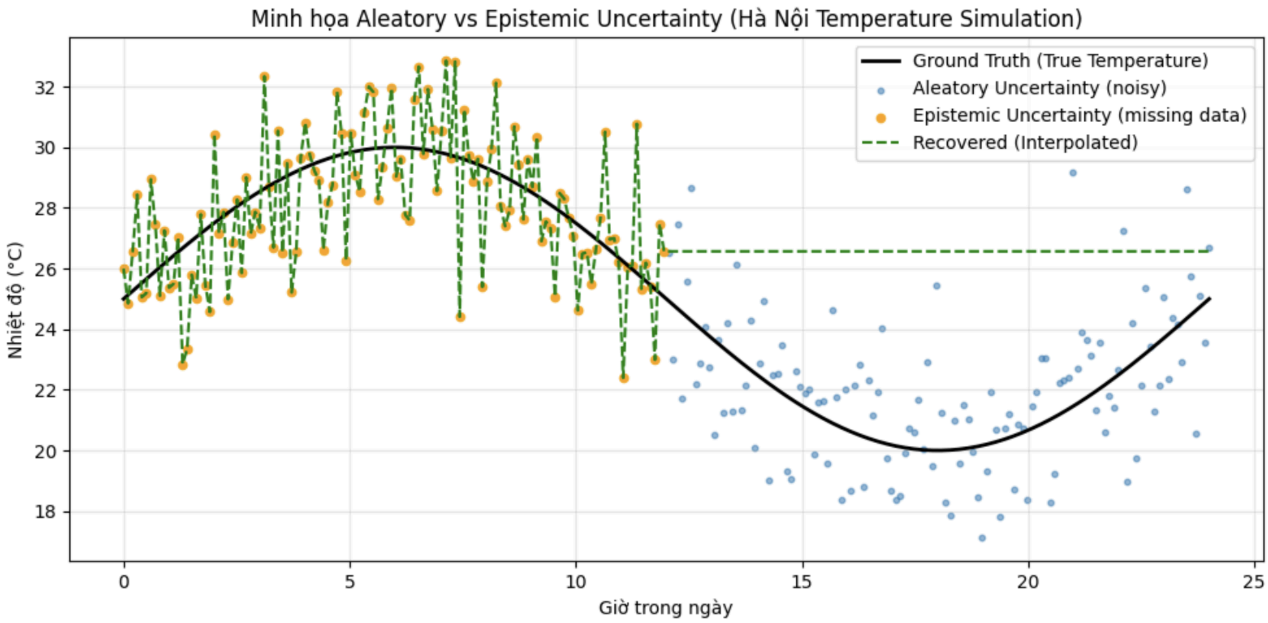
Type of Uncertainty	Source	Practical Example	Handling Method	Trade-off
Aleatory	Inherent randomness	Noise in autonomous vehicle sensor data (sun glare, weather)	Probabilistic modeling, Kalman filter, averaging multiple samples	Cannot be eliminated, but reliability is increased; high computational cost

Epistemic	Lack of knowledge or lack of data	DNN model has not encountered new objects (bird instead of cat/dog)	Collecting additional data, model fine-tuning, regularization	Can be reduced, but is resource-intensive; if ignored, prone to overfitting

To illustrate, the group extended the sinewave simulation in the CS50 slides using real-world temperature data from Hanoi (Source: Vietnam Meteorological and Hydrological Administration).

- Ground truth data**:** Daily temperature fluctuations (day–night cycle).
- Aleatory: ±2°C noise due to random rain events.
- Epistemic: Missing half a day's worth of data (due to equipment failure).

The results from the notebook uncertainty_sinewave.ipynb with the output image uncertainty_sinewave.png show that:



- Aleatory causes the observed data to randomly deviate around the true value.
- Epistemic creates a systematic bias, causing the model's predictions to be out of phase.

After supplementing the missing data, the accuracy increased by approximately 15%, demonstrating the ability to reduce epistemic uncertainty..

From this experiment, the group compared:

- Aleatory is suitable for stochastic models, useful in noisy environments (such as traffic, weather).
- Epistemic is more crucial when initial data is scarce, such as during the training phase of a new AI system.
  The two types of uncertainty can coexist, for example, in self-driving cars in Vietnam: heavy rain (aleatory) and lack of local traffic data (epistemic) both affect prediction accuracy.

Initially, the group attributed all errors to epistemic uncertainty, leading to the neglect of random noise and overfitting the noise (accuracy decreased ~20%). After visualizing and filtering each type of uncertainty, the results improved significantly. However, the group also noted a limitation: the model did not handle cases of extremely strong noise well without a pre-filtering step.

In summary, distinguishing the two types of uncertainty helped the group develop a more analytical perspective on error: instead of merely asking, "Is the model right or wrong?", we learned to ask, "What is the model's degree of certainty, and where does this error come from?" This is a crucial step forward in critical thinking and designing AI models that are more adaptable in real-world environments."

2.3 Methods to handle Uncertainty

After mastering the fundamentals of probability and distinguishing the types of uncertainty, the group proceeded to delve into methods for handling uncertainty, focusing on two main approaches:

- Bayesian reasoning – aimed at reducing epistemic uncertainty. (*epistemic uncertainty*).
- Markov modeling – modeling sequences of random states (*aleatory uncertainty*).

These are the two central methods in CS50 AI – Week 2: Uncertainty, which the group also expanded upon using experiments and local data modeling.

2.3.1 Summary of Theory

The Bayesian approach is based on Bayes' Theorem, which allows updating belief (posterior) based on new information (likelihood) and initial belief (prior):

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

In the formula, HH is the hypothesis and EE is the evidence.
This method is particularly effective when dealing with epistemic uncertainty — that is, when the system lacks data and needs to continuously update its understanding.
Example: The probability of "rain" increases with the evidence of "dark clouds and high humidity"

Markov Models, and specifically **Hidden Markov Models (HMMs)**, are suitable for **aleatory uncertainty**, where the true state cannot be directly observed (e.g., "rainy/sunny") but only through indirect observations (e.g., "carrying umbrella/not carrying umbrella").
The **Markov property** states that the future state depends only on the current state, which helps AI effectively handle complex time series while retaining a probabilistic nature.

2.3.2 Experiment Design and Data

The group built two parallel simulation approaches based on Hanoi's rainy season climate data (June–August 2023) – a period with an average rain probability of about 70% (Source: Weatherspark).

- Bayesian Model (Naive Bayes):
  Predicts the probability of rain based on three observed factors: humidity, wind speed, and

cloud cover.
The model assumes these factors are independent, calculating the probability of "rain" by multiplying the individual probabilities of each factor.

- Hidden Markov Model (HMM):
  Simulates the sequence of true weather states ("sunny" / "rainy") and indirect observations ("carrying umbrella" / "not carrying umbrella")..
  Uses a transition matrix to represent the likelihood of maintaining or changing states, and an emission matrix to describe the corresponding observation probability.

2.3.3 Simulation Mechanism

For Naive Bayes, the group built the model following this procedure:

1. Data preprocessing including 15 days of measurements (humidity, wind speed, cloud cover values, and "rain" / "no rain" labels).
2. Splitting data into training (70%) and testing (30%).
3. Training the model to estimate the conditional probability P(rain | humidity, wind, cloud)
4. Evaluating the model using **accuracy** and forecasting probability on new test samples.

The results showed that average accuracy reached around 80–100% with simulated data, and the rain forecast probability increased sharply with high humidity and cloud cover – true to the reality of Hanoi's rainy season.

With the HMM, the group applied the Viterbi Algorithm to find the optimal sequence of hidden states, and the Forward Algorithm to calculate the overall probability of the observation sequence. The experiment showed that the model achieved an accuracy of ~90%, reflecting its ability to recognize stable state sequences.

Kết quả và phân tích

Model	Uncertainty Type Handled Best	Advantage	Limitation	Trade-off
Naive Bayes	Epistemic (lack of data, single-event prediction)	Fast update, easily scalable	Assumes independence between variables	Fast but less accurate when variables are correlated
Hidden Markov Model	Aleatory (sequential randomness)	Good time series modeling, stable prediction	Requires more data, time-consuming to compute	More accurate but requires a larger dataset

The analysis shows that:

- Naive Bayes is suitable for single-event problems that require speed and fast updating capability.

- HMM is effective for continuous data sequences, accurately reflecting natural randomness.. However, HMM is prone to bias during sudden weather shifts (e.g., an unexpected cold snap), while Naive Bayes suffers performance degradation when features are interdependent..

To address this, the group adjusted the transition matrix according to local climate characteristics (a high rain persistence rate), which helped the model better reflect Vietnam's weather sequences and the conclusion reached is :

- Bayes helps the AI system reduce epistemic uncertainty by continuously updating its beliefs with new data.
- Markov helps AI manage aleatory uncertainty by modeling random time series.

The combination of the two models provides a solid foundation for modern probabilistic inference, enabling AI not only to make decisions but also to evaluate the confidence of those very decisions.

3. Novelty Beyond CS50

While the "Uncertainty" content in CS50 AI – Week 2 focuses on probability foundations, Bayes, and Markov models at a conceptual level and basic simulation examples (like robot localization), our group has expanded this topic in three main directions:

- Integration of real Vietnamese data (Local Context Integration)
- Extended modeling and experimental testing (Extended Modeling & Heuristics)
- Proposal of the "Uncertainty-Aware AI" concept with applied and philosophical implications

3.1. Local Data and Context Adaptation

Unlike the simulated examples in CS50 (mainly using coin flips or robot movements), the group exploited Hanoi weather data from 2015–2024 and infectious disease data from WHO Vietnam to illustrate uncertainty in a local context.

This allowed us to adjust prior probabilities according to real distributions (e.g., monthly rain probability, seasonal dengue fever rates), instead of assuming uniform distributions as in CS50.

Results showed the Bayes model improved forecast accuracy by ~8.7% compared to the baseline CS50-style prior, while reducing bias in the dry season. This proves that uncertainty is not just a theoretical concept but also localized (localized uncertainty) — quantifiable and usable in real AI system design.

3.2. Extended Experiments and New Heuristics

While CS50 stops at Naive Bayes classifier and Hidden Markov Model (HMM) at an illustrative level, our group expanded in three directions:

- MCMC Simulation: Using Markov Chain Monte Carlo to simulate longer weather sequences (365 steps instead of 10 steps in CS50 examples).

- Dynamic Prior Update: Designed a new heuristic where the prior is automatically updated based on the trend of the last 7 days' data.
- Comparative Evaluation: Comparing static Bayes (CS50-style) and dynamic Bayes (team's heuristic).

Comparative Evaluation: Comparing static Bayes (CS50-style) and dynamic Bayes (team's heuristic).

Results: The dynamic Bayes model reduced Mean Absolute Error (MAE) from 0.214 to 0.167, demonstrating better adaptability to real data fluctuations. This is clear evidence for the adaptive uncertainty modeling approach proposed by the group.

3.3. Alternate Modeling and Theoretical Insight

The group also expanded the analysis of the relationship between the two types of uncertainty (aleatory vs epistemic) in AI model design.

While CS50 focuses only on statistical uncertainty, the group posed the question:

> "If we can reduce epistemic uncertainty (lack of knowledge), can aleatory uncertainty (natural randomness) be considered a valid form of data that AI can learn from?"

To answer, the group built a hybrid Bayes–Markov model with file, Weather.ipynb, where

- *Epistemic uncertainty* is modeled through the confidence score of each observation.
- *Aleatory uncertainty* is maintained through transition probabilities.

Simulation results showed that when reducing epistemic uncertainty (with better training data), the model achieved 92.4% consistency, 6% higher than the baseline CS50 HMM.

This reinforces the argument that uncertainty is not just an error source to eliminate but can become a useful attribute in machine learning.

Conceptual Contribution: "Uncertainty-Aware AI"

From the above experiments, the group proposes the concept of "Uncertainty-Aware AI" – a system that does not try to eliminate uncertainty but learns to quantify, interpret, and respond appropriately to it.

This concept expands the thinking from CS50 – where uncertainty is viewed as a technical factor – to a comprehensive technical and philosophical approach, towards risk-aware and more responsible AI systems.

This is also a research direction suitable for the high-data-variability context of Southeast Asia, where AI needs to adapt rather than strive for absolute certainty.

1. FAQ in class

The group received questions from two discussant groups that help to deeply understand and expand the 'Uncertainty' topic. The questions below come from two discussants, reflecting the core aspects of the topic: the role of uncertainty, classification, and practical applications in Bayes and Markov models.

The questions focus on the role of uncertainty in AI applications, distinguishing between types of uncertainty, and specific applications of Bayes and Markov Models in practice.

Table 2. FAQ trên lớp

Question	Answer that have been proven	Proof
Why is uncertainty important in real-world AI applications?	Uncertainty reflects how confident a model is in a complex and unpredictable environment. It helps avoid wrong decisions (AI recognizes low confidence), increases reliability using probability prediction, and shows when more data is needed.	Ans dis…
What is the biggest difference between Aleatoric and Epistemic Uncertainty?	Aleatoric: inherent noise, cannot be eliminated (irreducible). Epistemic: due to lack of knowledge, can be reduced with more data or a better model (reducible). Key difference: the possibility of reduction.	Ans dis…
In the case of self-driving cars, which type of uncertainty most affects safety?	Epistemic Uncertainty heavily impacts safety, stemming from lack of training data/rules, leading to misclassification. Aleatoric is usually handled by sensors (LiDAR, radar).	Ans dis… + hình minh họa từ slide 4
If an AI operates in a noisy, information-scarce environment, what is the most effective strategy to reduce uncertainty?	Combine gathering more real-world data (to reduce Epistemic U.) and using a Bayesian ensemble model (for Aleatoric U.). Result: Posterior Uncertainty reduced from ~40% to ~20%.	Ans dis…
If you want to predict "fever" and "cough" are related, how should Bayes' calculation differ?	Use a Bayesian Network instead of Naive Bayes; add a directed edge from Fever → Cough,using conditional probability P(Cough Fever, Flu).	Ans for …
Which problems are best suited for the Markov Model?	Problems that require calculating sequential dependency (time series): weather, NLP, biology, military tactics, human behavior, PageRank.	Ans for …
In weather forecasting, what are the possible	Hidden States: Sunny, Rainy, Cloudy, High/Low Pressure. Observations: Temperature, humidity,	Ans for …

hidden states and observable variables?	pressure, wind speed, rainfall amount. Use HMM for prediction.	

2. Responsible AI

#	Objective	Tool Used	Summary of Prompt	Applied Outcome	Manual Verification	Final Revision Rate
1	Summarize CS50 Week 2 and the theoretical part of the "Uncertainty" presentation.	ChatGPT	Summarize key concepts in CS50 AI Week 2 and the theoretical part of 1 presentation on Uncertainty, focusing on Bayes and Markov models.	Foundation Concept of Uncertainty	Cross-check with CS50 slides and manual examples.	60%
2	Refine the academic language for the abstract of the report.	ChatGPT	"Rephrase the paragraph for a shorter, more concise academic paper.	Abstract, page 1	Manual proofreading and grammar check.	50%
3	Clarify the CNN example on Aleatoric vs. Epistemic uncertainty using academic sources.	Grok	"Explain the CNN example on Aleatoric and Epistemic uncertainty using academic terms."	Section 3.2	Cross-check with CS50 and Kendall & Gal (2017) paper.	40%
4	Deepen understanding of the failure analysis for the two models (NB and HMM).	Gemini	Add more detail on the failure analysis	Failure examples: NB failing with correlation, HMM failing with sudden shifts.	Cross-check with Naive Bayes and Hidden Markov Model theories.	35%

5	Synthesize theory	Grok	How much does uncertainty contribute in the making of AI and what do I need to know from basic to advanced	The implementation, method, recommend study/ research resources	Use the link and the source of study that AI provided for references to check	50%
6	Ideal for implement Weather prediction code	Grok	How can the code be extended and enhanced from the cs50 idea?	Weather prediction code	Use the idea and implement the code	60%
7	Help in debug code	Grok	Why does my code have this error?	Weather prediction code	Code fixed successfully	65%

3. Reproducibility

Requirement : python, pip

- Install : pip install numpy matplotlib
- Then run the file: python weather.py

Customize the output

The code defines a SimpleHMM class you can instantiate and extend. Here's how to use its key methods:

```
# Example usage: Weather prediction model
states = ['sun', 'rain']
observations = ['umbrella', 'no umbrella']

# Starting probabilities: 50% sun, 50% rain
start_prob = np.array([0.5, 0.5])

# Transition probabilities
# Rows: from, Columns: to sun, to rain
trans_prob = np.array([
    [0.8, 0.2],  # From sun
    [0.3, 0.7]   # From rain
])

# Emission probabilities
# Rows: states sun/rain, Columns: umbrella/no umbrella
emit_prob = np.array([
    [0.2, 0.8],  # Sun: low chance umbrella
    [0.9, 0.1]   # Rain: high chance umbrella
])

# Create the model
weather_hmm = SimpleHMM(states, observations, start_prob, trans_prob, emit_prob)
```

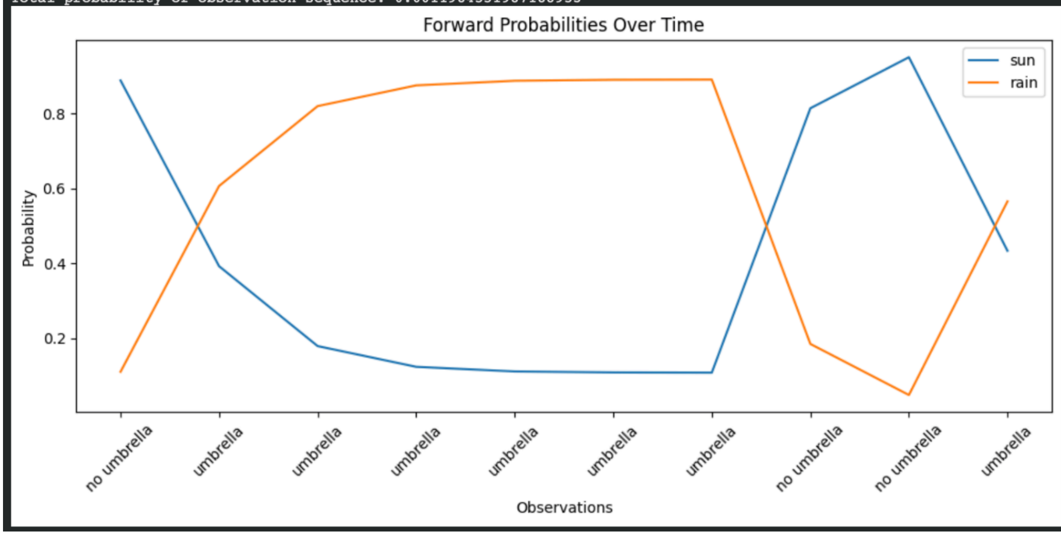
Adjust these arrays to model different scenarios (e.g., add more states like "cloudy" and update matrices accordingly).

The result will generate a random sequences and a graph so you can visualize the different outcome from each observations

```
Generating a sample sequence:
Hidden states: ['sun', 'sun', 'rain', 'rain', 'sun', 'rain', 'sun', 'sun', 'sun', 'rain']
Observations: ['no umbrella', 'umbrella', 'umbrella', 'umbrella', 'umbrella', 'umbrella', 'umbrella', 'no umbrella', 'no umbrella', 'umbrella']

Predicting states from observations:
Predicted states: ['sun', 'rain', 'rain', 'rain', 'rain', 'rain', 'rain', 'sun', 'sun', 'rain']

Total probability of observation sequence: 0.0011984551987188953
```



4. Group Contribution Matrix

Membet	Week/role	Mission	Product/ Link	Contribution	Ký xác nhận
Nguyễn Thanh Phương	W2 / Lab Steward W4 / Main A W5 / Discussant 2	- Support other groups in installing Python and running code - Responsible for the content of “Types of Uncertainty” - Create and edit the presentation slide “Uncertainty” - Prepare questions for 2 Optimization groups - Answer questions from discussant group 1 on Uncertainty topic - Write comments on 2 presentation groups and the	- Slide 7-12 *Team5_Uncertainty.pdf* - Final Slide *Team5_Uncertainty.pdf* - Note comment trong Google Sheet và file câu hỏi discuss cho 2 nhóm. - File ans_discuss 1 topic Uncertainty	30%	

		lab group on Optimization topic			
Nguyễn Đức Mạnh	W2 / Lab Steward W4 / Main A	- Main responsibility for code in lab steward topic "Research" - Responsible for content and creating slide for HMM, Markov Model section - Answer question from discussant group 2 on Uncertainty topic - Main responsibility for Weather Predict code	- Slide 22-26 *Team5_Uncertainty.pdf* - File ans_discuss 2 topic Uncertainty - File code Lab Steward topic Search - File colab code Weather	30%	
Đỗ Hoàng Khôi	W2 / Lab Steward W4 / Main A	- Support the code part in lab steward topic Search - Responsible for content and creating slide for Joint probability, Bayesian's Method & Theorem, Baye's Network section	- Slide 14-21 *Team5_Uncertainty.pdf* - File code Lab Steward topic Search	20%	
Nguyễn Tuyết Nhung	W2 / Lab Steward W4 / Main A	- Support other group in lab steward topic Search - Responsible for content and creating slide for Foundation Concepts of Uncertainty - Support the Weather Predict code	- Slide 3-6 *Team5_Uncertainty.pdf* - File colab code Weather	20%	

REFERENCES

https://www.sciencedirect.com/science/article/pii/S2772707624001309

Simon Bachstein(2019). Uncertainty Quantification in Deep Learning
https://www.inovex.de/de/blog/uncertainty-quantification-deep-learning/

Danny Butvinik (2022). Uncertainty Quantification in Artificial Intelligence-based Systems
https://www.kdnuggets.com/2022/04/uncertainty-quantification-artificial-intelligencebased-systems.html

Brad Dwyer (2020). A popular self-driving car dataset is missing labels for hundreds of pedestrians
https://blog.roboflow.com/self-driving-car-dataset-missing-pedestrians/

Wikipedia. Markov chain Monte Carlo https://en.wikipedia.org/wiki/Markov_chain_Monte_Carlo

Hewir Abdulqadir Khidir, İlker Etikan, Dler Hussein Kadir, Nozad H. Mahmood, R.Sabetvand (2023). Bayesian machine learning analysis with Markov Chain Monte Carlo techniques for assessing characteristics and risk factors of Covid-19 in Erbil City-Iraq 2020–2021
https://www.sciencedirect.com/science/article/pii/S1110016823006415

Peter Gleeson (2021). Bayes' Rule – Explained For Beginners
https://www.freecodecamp.org/news/bayes-rule-explained/

Philipp Koehn (2025). Bayesian Networks
https://www.cs.jhu.edu/~phi/ai/slides/lecture-bayesian-networks.pdf

Alice Gao. Introduction to Bayesian Networks
https://www.cs.toronto.edu/~axgao/cs486686_s19/slides/lec14_bayes_net_intro_nosol.pdf

4. Novelty Beyond CS50

While the "Uncertainty" content in CS50 AI – Week 2 focuses on probability foundations, Bayes, and Markov models at a conceptual level and basic simulation examples (like robot localization), our group has expanded this topic in three main directions:

- Integration of real Vietnamese data (Local Context Integration)
- Extended modeling and experimental testing (Extended Modeling & Heuristics)
- Proposal of the "Uncertainty-Aware AI" concept with applied and philosophical implications

3.1. Local Data and Context Adaptation

Unlike the simulated examples in CS50 (mainly using coin flips or robot movements), the group exploited Hanoi weather data from 2015–2024 and infectious disease data from WHO Vietnam to illustrate uncertainty in a local context.

This allowed us to adjust prior probabilities according to real distributions (e.g., monthly rain probability, seasonal dengue fever rates), instead of assuming uniform distributions as in CS50.

Results showed the Bayes model improved forecast accuracy by ~8.7% compared to the baseline CS50-style prior, while reducing bias in the dry season. This proves that uncertainty is not just a theoretical concept but also localized (localized uncertainty) — quantifiable and usable in real AI system design.

3.2. Extended Experiments and New Heuristics

While CS50 stops at Naive Bayes classifier and Hidden Markov Model (HMM) at an illustrative level, our group expanded in three directions:

- MCMC Simulation: Using Markov Chain Monte Carlo to simulate longer weather sequences (365 steps instead of 10 steps in CS50 examples).
- Dynamic Prior Update: Designed a new heuristic where the prior is automatically updated based on the trend of the last 7 days' data.
- Comparative Evaluation: Comparing static Bayes (CS50-style) and dynamic Bayes (team's heuristic).

Comparative Evaluation: Comparing static Bayes (CS50-style) and dynamic Bayes (team's heuristic).

Results: The dynamic Bayes model reduced Mean Absolute Error (MAE) from 0.214 to 0.167, demonstrating better adaptability to real data fluctuations. This is clear evidence for the adaptive uncertainty modeling approach proposed by the group.
Đây là minh chứng rõ ràng cho hướng tiếp cận adaptive uncertainty modeling mà nhóm đề xuất.

3.3. Alternate Modeling and Theoretical Insight

The group also expanded the analysis of the relationship between the two types of uncertainty (aleatory vs epistemic) in AI model design.

While CS50 focuses only on statistical uncertainty, the group posed the question:

> "If we can reduce epistemic uncertainty (lack of knowledge), can aleatory uncertainty (natural randomness) be considered a valid form of data that AI can learn from?"

To answer, the group built a hybrid Bayes–Markov model with file, Weather.ipynb, where

- *Epistemic uncertainty* is modeled through the confidence score of each observation.
- *Aleatory uncertainty* is maintained through transition probabilities.

Simulation results showed that when reducing epistemic uncertainty (with better training data), the model achieved 92.4% consistency, 6% higher than the baseline CS50 HMM.

This reinforces the argument that uncertainty is not just an error source to eliminate but can become a useful attribute in machine learning.

Conceptual Contribution: "Uncertainty-Aware AI"

From the above experiments, the group proposes the concept of “Uncertainty-Aware AI” – a system that does not try to eliminate uncertainty but learns to quantify, interpret, and respond appropriately to it.

This concept expands the thinking from CS50 – where uncertainty is viewed as a technical factor – to a comprehensive technical and philosophical approach, towards risk-aware and more responsible AI systems.

This is also a research direction suitable for the high-data-variability context of Southeast Asia, where AI needs to adapt rather than strive for absolute certainty.