**Speaking suggestions:**

QEQR

Adrenaline rush before going onto stage.

Nervous? Speak slowly.

Nervousness = you are about to do something exciting.

People won’t even remember that you were there.

Use pauses to lay stress.

Smile and use open hand gestures.

Have melody in voice.

Try to start with some attention grabber.

Recall answers by other panellists, it makes you more memorable.

Immediate TODO: video recordings, and speak 20 pc louder than.

**Generative AI -**

Existing Legislations on protection of customer data - GDPR (EU) and FCRA (US), plus more are being formed for accountability, transparency, and risk management.

For e.g. Recently launched (1 Aug 2024) European Union's AI Act proposes stringent transparency obligations for high-risk AI systems.

So, one must be mindful of these before building any solution for the customer.

Hence despite pressure on management in some cases to try out gen ai as a use case, banks are and should take time to adopt gen ai to see how the regulations develop first.

GDPR (General Data Protection Regulation) mandates appointment of DPO (Data Protection Officer) when processing large scale of personal data for the AI solution, failing which it can lead to heavy fines. GDPR also mandates taking customer's consent on use of their sensitive data like health, biometrics, ethnicity etc.

Another act to be mindful is of DORA (Digital Operational Resilience Act) - another European regulation effective from Jan 2025 - Banks should be able to withstand and recover from severe operational disruptions, which includes 3rd party solutions being used. So, if using a 3rd party solution for generative AI use case, be mindful of this act.

Banks now usually do not own high GPU chips, instead leverage cloud offerings.

Human intervention is required even with the use of Gen AI. e.g. technique called RLHF - Reinforcement learning from human feedback, use humans to evaluate the fine tuning of the model.

Common problems with AI systems: Bias, Overfitting, generating harmful content etc.

Gen AI problems: Hallucinations, harmful content as source is unknown, Prompt Injection - info leakage.

Use Gen AI as more of a phone a friend kind of help, do not trust it blindly.

----

NatWest Group is leveraging generative AI in following ways:

1) Collaboration with AWS: NatWest has teamed up with Amazon Web Services (AWS) to develop AI products by 2027 to help customers with their financial planning, conduct Financial Health Check and check credit score.

2) Cora+ Digital Assistant: NatWest has upgraded its digital assistant, Cora, to Cora+, which uses generative AI to provide more intuitive and conversational customer support. This upgrade, developed in collaboration with IBM, allows Cora+ to offer proactive assistance and more personalized responses based on past interactions and transaction history.

3) Fraud Detection: By analysing customer behaviours with AI, NatWest can detect unusual payment patterns earlier, helping to prevent scams and reduce financial loss.

----

As per European AI Act, there are 4 categories of risk when using AI enabled systems:

Cat 1: Minimal Risk. No regulatory intervention. E.g. ai enabled video games or spam filters.

Cat 2: Limited risk. Key is 'Transparency': User just need to be informed that they are dealing with AI generated content. e.g. deepfakes, ai generated text/images and chatbots.

Cat 3: High Risk. E.g. Employment decisions, healthcare, transport (self-driving cars), Education, Law enforcement etc. Key for this category is to go under 'rigorous risk assessment, quality check for datasets, maintain activity logs, and comprehensive documentation.'

Cat 4: Unacceptable risk: e.g. real time facial recognition in public for law enforcement purposes, **social scoring** - individuals are ranked based on certain behaviours or characteristics. e.g. China does it to score what loans individuals can apply etc. This type of use case is strictly banned in EU.

ChatGPT falls in cat 2, but new regulations are being laid out for companies to share the data used for training the models. This would enable the authors of content seek compensation, which the tech companies that have built these models are strictly against.

Act is already in force now from 1 Aug 2024, but rules will be gradually applied. Companies have 2 years to comply with the act, but only 6 months to remove Cat 4 forbidden use cases. Big tech companies like Microsoft, Google, and Amazon have already accepted the relation and adapting to the same.

A Fundamental Rights Impact Assessment (FRIA) is a key component of the EU AI Act, particularly for high-risk AI systems. The FRIA aims to evaluate and mitigate potential risks that an AI system might pose to fundamental rights of individuals.

Act says that use case defines the risk. For e.g. if you use a chat solution to ask recommend best restaurants in xyz area, this is very innocuous. But if you ask the app which of these restaurants, should I invest my money on, then suddenly this becomes a high-risk use case, and the developer of the app must adhere to the regulation.

EU Act is applicable to UK as well as most of the solutions developed will be accessible in EU market, hence applicable for EU act.

UK govt however is working on its own regulation as well, which is expected to be a bit less strict and less prescriptive than EU act.

Certain AI use cases in Banking can be as 'high risk' like checking credit worthiness of an individual and risk assessment in life and health insurances, fraud detection and anti-money laundering.

The act defines 4 entities and introduces specific roles and responsibilities for them - 'providers' (developers, the entity that develops the ai solution) and 'deployers' (users, the entity using AI system under its authority). There are Importer and Distributor as well, but provider and deployers are the main ones. Provider need to ensure that they follow Act's risk assessment and adhere to guidelines. Deployers of high-risk solution must follow provider's instructions, ensure human oversight, monitor systems operations, and report any risks to the provider or relevant authorities.

Most responsibility of adhering to the act is on developers, and some on deployers.

If you develop AI solution in house, then you are both the provider and the deployer of the system.

Human oversight means that in case AI systems misfire, then we can have human intervention to take care of the situation.

So we see that EU AI act not just defines horizontal risk categories, but also kind of vertical responsibilities in terms of providers/deployers, so a matrix kind of responsibility when building ai systems.

-----------------------------------------------------------------------------------------------------------------------------------

Key principles across legislations: Fairness, Transparency, Accountability, Human rights, Lawfulness.

**Ethical AI is about aligning use of AI to societal values.** It is not just good practice but good business. Idea is not to restrict use of AI, but to get full benefits of AI in right ways.Note that Generative AI is still very young and growing technology in terms of adoption. Not many examples of AI ethics gone wrong are there, but in due time there will be. Better be prepared now, than be sorry later**.** Adopt AI ethics from the very beginning of your solution.

**Principles of AI Ethics:  
1. Fairness:** AI solution should generate results equal to all, without any **bias**. Gen AI models are trained on data created by humans only. Hence the biases can reflect in the model responses as well. Eg. It can lead to insurance claim of entire neighbourhood getting rejected. Or a bank may tend towards higher income customers when using AI enabled systems. Solution is to re-engineer your prompts, fine tune the models where u see a bias, and do proper testing with wide variety of data to confirm there is no bias.

2. **Accountability**: clear responsibilities defined for each of AI system’s outcome. Robust mechanisms to recover from any failure or deviation. Both the producers and the consumers are accountable.

3. **Transparency**: make AI systems explainable and understandable. Share what data has been used, how it was processed and what models have been used. Use techniques like XAI.

**Commitment**: Adherence to above principles builds trust and mitigates risk.  
----------------

**Privacy-Personalisation Paradox**: To give personalised recommendations, access to personal data is required. But how much access should be allowed? This is known as privacy personalization paradox.

Solution is **AI literacy**: inform the users how their data is being utilised.  
  
Another challenge is that highly accurate models are complex and are less transparent as compared to simple not so accurate models. This is also where AI literacy comes to rescue.

If you really must train model on sensitive customer data, remember to anonymize customer data first.  
---------------------------------------------

**Bias-Fairness conundrum**: Underlying training data is human created only, hence there can be an inherent bias in the solution built.

AI should not be used as the final decision maker. It should only aid decision making. There should be a human involvement at some step.

**Paradox of Accountability**: The paradox of accountability with generative AI revolves around the tension between the impressive capabilities of these models and the challenges in holding them accountable for their outputs. Getting too creative can lead to wrong outputs.

**Explainable AI** (**XAI**) refers to a set of processes and methods that make the outputs of machine learning models understandable to humans. 2 popular XAI techniques:  
**LIME (Local Interpretable Model-agnostic Explanations)**

It acts a translator to explain functioning of the model. LIME explains individual predictions by approximating the complex model with a simpler, interpretable model around the instance being explained.

**SHAP (SHapley Additive exPlanations)**

Explains what features of data set had more weight in decision making. Eg. The director of the move had 50pc weight in deciding it as a recommended movie.

Assurance mechanisms:

Choose your base model wisely, wrt regulations.

Educate and train your employees.

Train your model, and test it properly.

Inform your customers about use of AI in decision making. Be transparent.

Proper monitoring and auditing post model deployment.

Automated processes to have human in the loop for final decision making.

Do due diligence to ensure data protection, fairness assessment.

Consumer LLMs vs. Enterprise LLMs (trained from scratch internally), but banks usually don’t own such large GPU machines, and cost is also high for training from scratch on cloud.

**PANEL DISCUSSION**

**1.Understanding Generative AI**

o Definition and applications in banking

A: Generative AI is a type of artificial intelligence (subset of DL) that can create new content such as text, images, videos, or audio based on patterns learned from existing data.   
You must be wondering – ‘how is gen ai different from existing AI models?’

The response generated by Gen AI models are ‘contextual’ i.e. preserves the meaning of user’s input.

o Examples of generative AI models

Gen AI models can be used to generate text, images, audio, videos.

For banking perspective, text is the commonly used gen ai type. Use cases types: chat solutions, content creation like underwritings and text summarisation.

Some popular text-based models: GPT-4 (by openai), Llama (metaAI), BERT, Cohere.

Q. What are some practical applications of generative AI in banking?

Ans.

1. **Lending** process: Automate lending process. Marketing and lead generations: chatbots for potential customers. Lending product recommendations based on customer profile.

2. **Loan Applications**: chatbots can assist clients in loan application from completion.

3. **Credit ratings**: Gauge credit worthiness of a client based on history and other parameters like gauging riskiness of a loan application.

4. **Load underwriting**: help underwriters churn through the data quickly.

5. **Account mgt**: Help customers with Account mgt, billing questions, and payment reminders. Based on payment history, give recommendations for debt payback.

6. Financial planning.

7. Fraud Detection

8. Algorithmic trading

(general company use cases):

6. **Employee productivity**: HR system chatbot, code generation

7. **CRM: Summarisation** of user requests/complaints

8. **Customer communication**: e.g. Emails in simple language for customers to understand financial terms.

Q. How do generative AI models differ from traditional AI models?

* **Traditional AI**: Designed to execute predefined tasks, such as classification, prediction, and automation. More algorithmic.
* **Generative AI**: Capable of creating new content. More creative as they preserve the ‘context’ of data learned and user’s input prompt. Something traditional AI is not capable of.

**2. Ethical Considerations**

o Fairness and bias in AI

o Transparency and explainability

Q. How can banks ensure fairness and reduce bias in AI models?

Ans. We all are familiar with a term ‘Fairness-bias conundrum’: Models are trained on human generated content. So Bias is there. You need to ensure your AI system is free of any bias for your use cases.

- Fine tune the model on diverse set of data.

- Detect bias using fairness metrics.

- Keep humans in the loop, detect any drift in model output.

- Transparency.

Q. What steps can be taken to make AI decisions more transparent?

Ans.

- End users must be clear about use of AI, and not a human generated output.

- Customer should be aware what kind of data is being used.

- Maintain documentation around models and data being used.

- Explainability: Use techniques like XAI (LIME/SHAP) to explain the model better.

- External audit and validation.

**3. Risk Management**

Q. Identifying and mitigating AI risks

Ans.

1. **Bias and fairness**: fine tune your model on a diverse set of data, and ensure wholistic testing.

2. **Hallucination**: Proper fine tuning, techniques like RAG can help reduce hallucination.

3. **Prompt injection**: sanitise user’s input, apply guardrails on the responses from your AI model.

4. **Data Leakage**: Apply guardrails. So both the input and output has to be sanitised.

Most important is to keep a human in the loop to take over should things go wrong.

Q. Strategies for robust AI governance

(Ensuring compliance with AI regulations)

**4. Compliance Requirements**

Q. Regulatory landscape for AI

Ans. AI, and generative AI is still a very new technology, and is evolving as being adopted. Hence the regulations are evolving as well.

At the forefront is EU AI Act, which defines 4 risk categories. The act also defines 2 important roles – ‘provider’, ‘deployer’ and defines different set of responsibilities for both. So you see that the act kind of defines both horizontal and vertical concerns one needs to take care of.

Also, there are other mandatory regulations like GDPR, SOX, DORA, and non-mandatory frameworks like NIST, ISO 27001 etc.

Since it is an evolving landscape, we will hear more about firms being fined for violating regulations in near future.

Q. Ensuring compliance with AI regulations?

Ans.

1. Define chain of accountability.
2. Protect customer data when fine tuning your models.
3. Transparency – document your solution, and use techniques like XAI.
4. Stay updated with regulatory changes and **continuously adapt** your AI systems.
5. Regular audits and monitoring.
6. Ensure that there is always some human in the loop to detect any drift.
7. Users of AI system should be aware of their side of responsibilities, and they should use the system ethically.

**5. Data Privacy and Security**

Q. Protecting customer data

Ans. First identify **sensitive** data, as this is the data that needs protection.

3 core pillars of data security;

Data confidentiality (encryption, deletion before disposing how off),

Data integrity (verify validate data, secure log changes to data),

Data availability (backups)

Data confidentiality – 4 pillars:

**Fundamental data protection** measures:

1. Encryption of sensitive data
2. Secure hardware
3. Role based access + Zero trust architecture

**Gen AI specific data protection**:

1. All personal data was anonymised when training the model.
2. Handle prompt injection attacks.
3. Apply guardrails to deny responses of model that contain sensitive/harmful content.

**Mandatory regulations**: legally binding.  
 GDPR (EU – protect customer data) grants individuals extensive right over their data.

SOX – regular financial reporting for more accountability and to prevent financial frauds.

CISA Act – US law that mandates sharing of cybersecurity threats by pvt. companies with federal govt.

**Voluntary frameworks**: not binding. Companies may use them to reduce their cybersecurity risk and increase their credibility. E.g.:

ISO 27001: Orgs can opt to get certified by choosing external audit.  
NIST CSF: Best practices and common language around cyber security.

COBIT: framework to minimize IT risk.)

Q. Addressing data breaches and leaks.

Ans.

1. Avoid further leak. Isolate the systems impacted and disable the gen AI solution.
2. Assess the impact of data breach.
3. Inform stakeholders and regulators. And based on consultation with them, inform the impacted customers.
4. RCA: Analyse logs and identify how the breach happened.
5. Roll out the fix with proper testing.
6. Adhere to regulatory guidelines, and opt for optional frameworks and audits to regain lost trust.

**Other Related Themes:**

• AI Auditing: Regular audits for AI systems to ensure compliance.

• Ethical AI Frameworks: Developing and implementing ethical guidelines for AI use.

• Stakeholder Engagement: Involving customers and regulators in AI discussions.