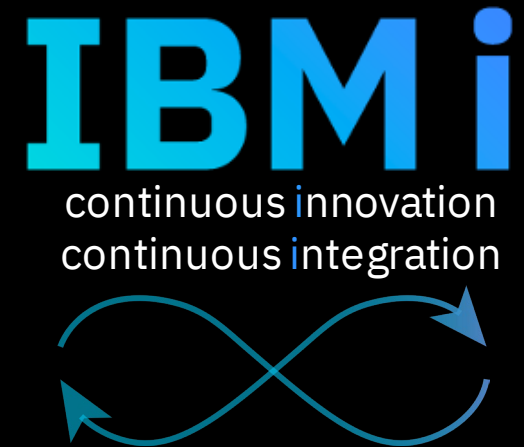


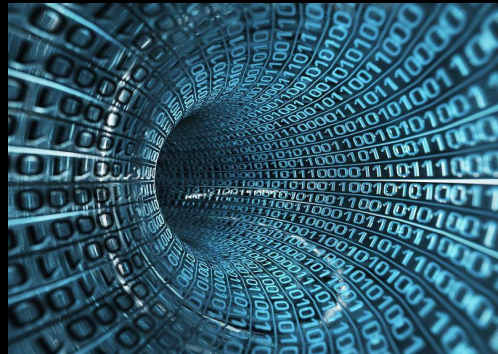
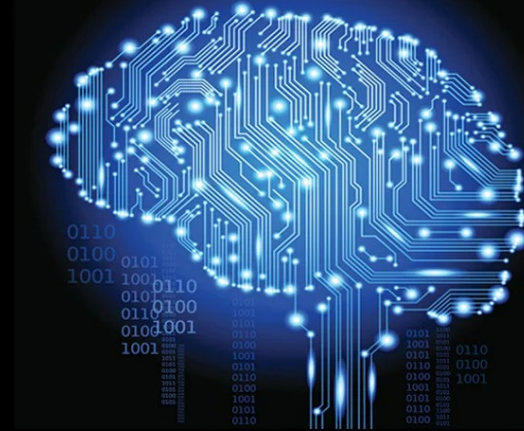
IBM i & AI – Strategy & Update

Steve Will – IBM i CTO & Chief Architect

July 2024

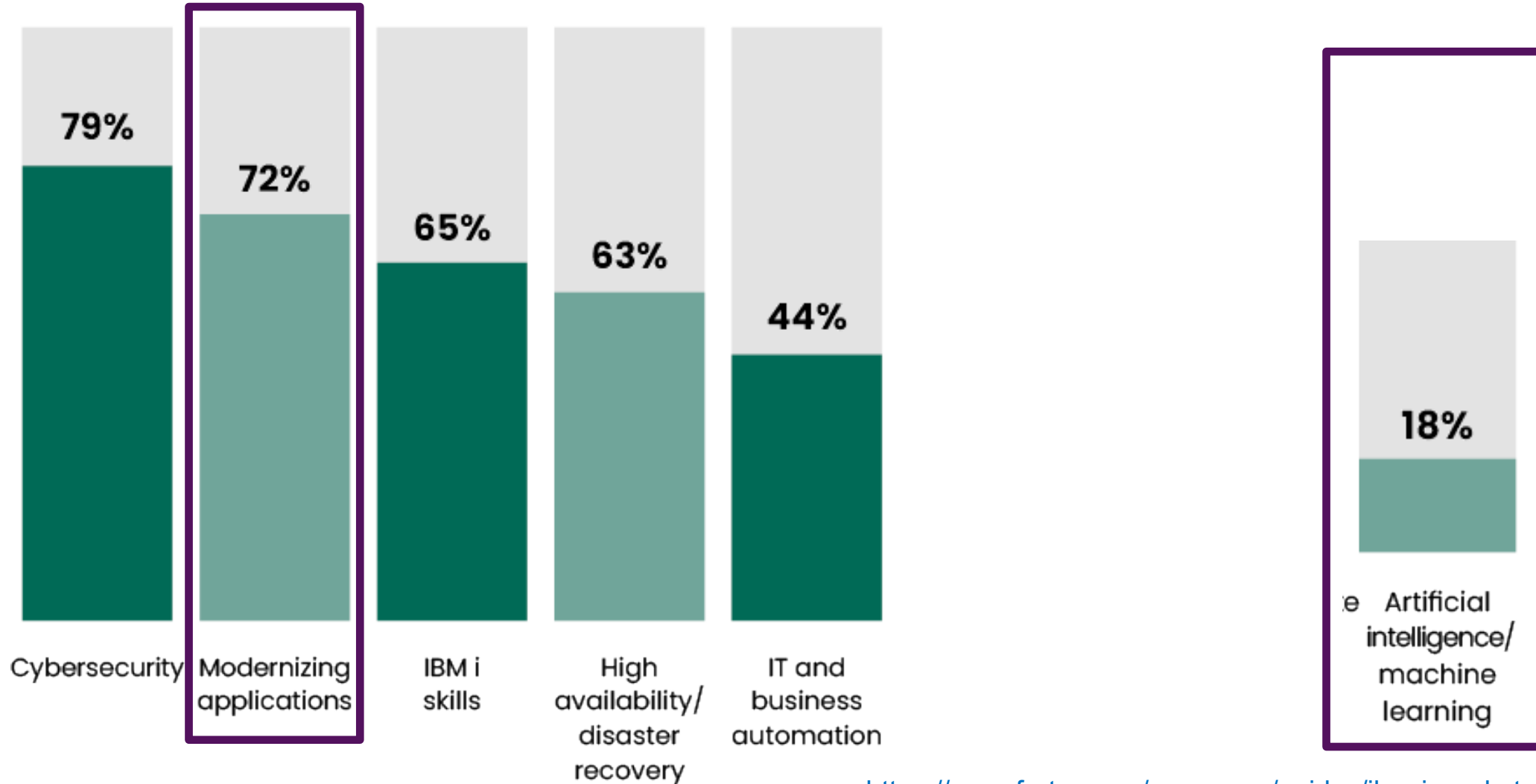


IBM i

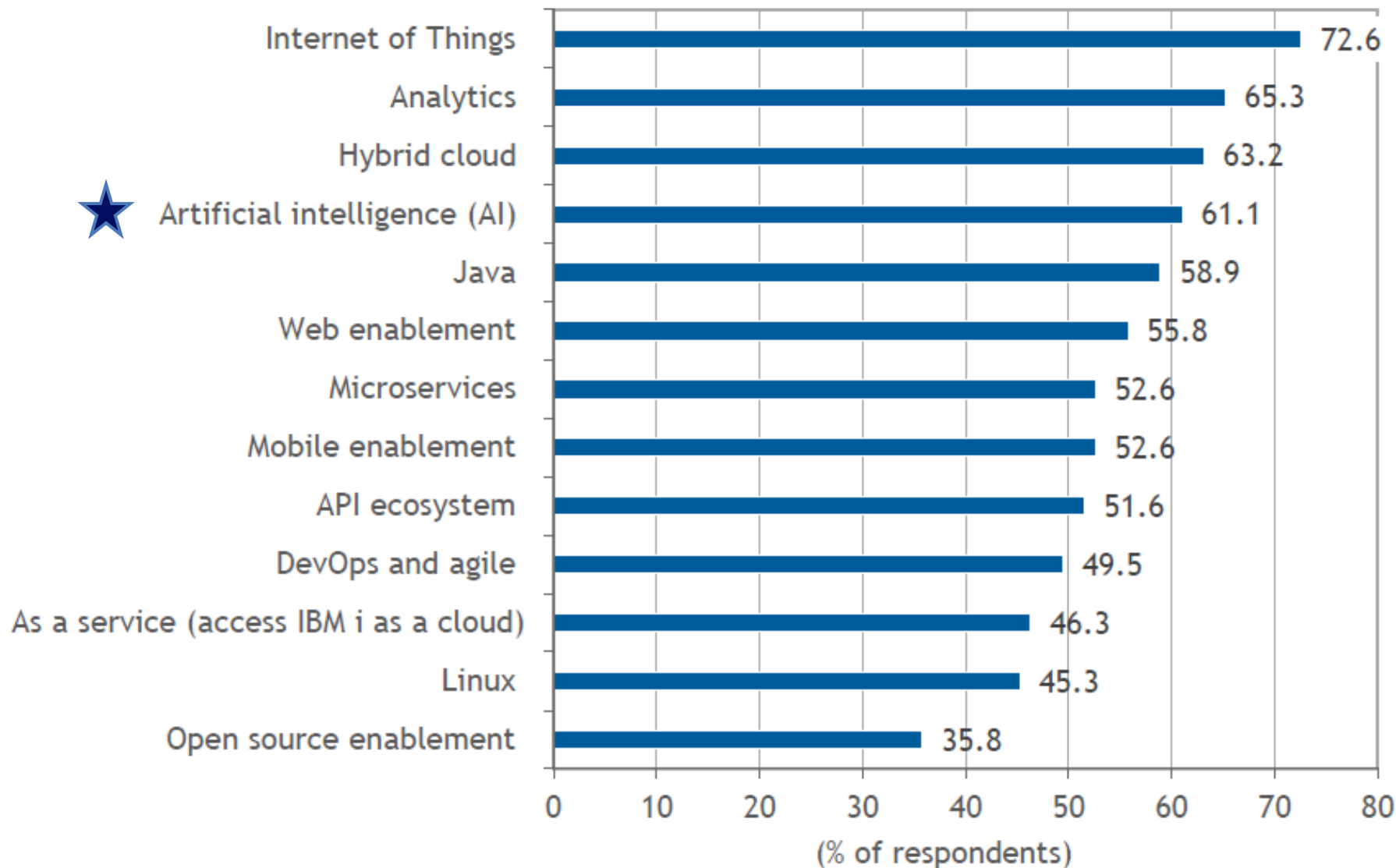


Top IT Concerns

What are your top 5 concerns as you plan your IT environment?



Modernization projects on IBM i – survey published 2020



Source: IDC, 2020

IBM i and AI – Historical & Current

- ~2016+ – Customers use IBM i Open Source to implement AI on IBM i
 - Credit Score validation and risk assessment
 - Training for Agricultural Business ERP
 - Tracking/Predicting food safety
 - Guiding/Predicting customer preferences pre-manufacturing
 - Chat-bots for operations and customer Q&A
- 2021+
 - Power10 Announce demo – IBM i workload integrating P10 MMA Linux use
 - Various client engagements to nurture and PoC AI enablement technology
- Watson Geospatial integrated into Db2 for i
 - Immediately picked up by several clients in logistics, agriculture, etc.
- IBM i team implementing various technologies
 - Extensions to existing VS Code based tools which can prompt a user's LLM
 - Underpinnings for strategy

IBM i Strategy



Power Solutions

- Enable clients to exploit latest Power technology
- Enable transformation of customer solutions with new value
 - Mobile, Internet of Things, Cognitive, Machine Learning and AI
- Enable Solutions to modernize around services, hybrid cloud and DevOps



Open Platform for Choice

- Grow IBM i solutions options including
 - Modern RPG & Open source technologies
- Flexible options
 - On-prem and/or Hybrid Cloud - inside or outside Data Center
- Entice new talent with popular open languages and tools



The *Integrated* Promise of IBM i

- Deliver a simple, high value platform for business applications
- Provide exceptional security and resiliency for critical business data
- Leverage IBM systems, storage and software technologies



IBM i Strategy



Power Solutions

AI that works with **their** solutions & data



Open Platform for Choice

AI options



The *Integrated* Promise of IBM i

AI that's easy to deploy & manage



IBM i + AI

Db2 Data Analytics

- Trend analysis
- Anomaly detection

Operations

- Active monitoring / alerting
- Self-healing

Developer Experience

- Help developer write code
- Understand code

Db2 Data Analytics

- Trend analysis
- Anomaly detection

Db2 Data Analytics

- Trend analysis
- Anomaly detection

Operations

- Active monitoring / alerting
- Self-healing

AI Workloads

AI Analytics / Fraud Detection

1. AnswerRocket
2. Cognos / IBM Watson Analytics
3. Power BI
4. Oracle Analytics Cloud
5. The Analytics OS (Pyramid v2020)
6. Qlik
7. Einstein Analytics Platform
8. SAP Analytics Cloud
9. SAS Visual Analytics
10. Sisense

Ref: <https://solutionsreview.com/business-intelligence/the-best-ai-analytics-software-tools/>

AI Ops

1. AppDynamics
2. DataDog
3. BigPanda
4. New Relic
5. Instana
6. Moogsoft
7. LogicMonitor
8. Splunk Enterprise
9. PagerDuty
10. BMC

Ref: <https://www.unite.ai/top-aiops-platforms-tools/>

Developer Experience

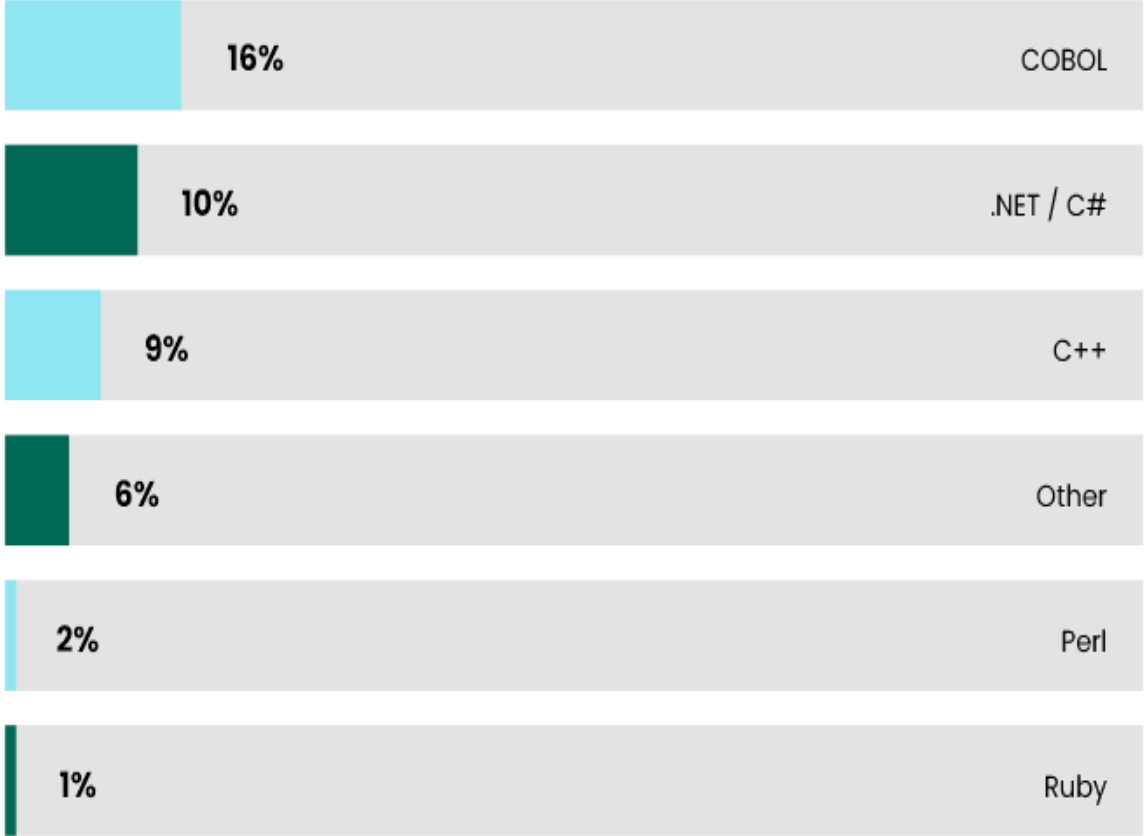
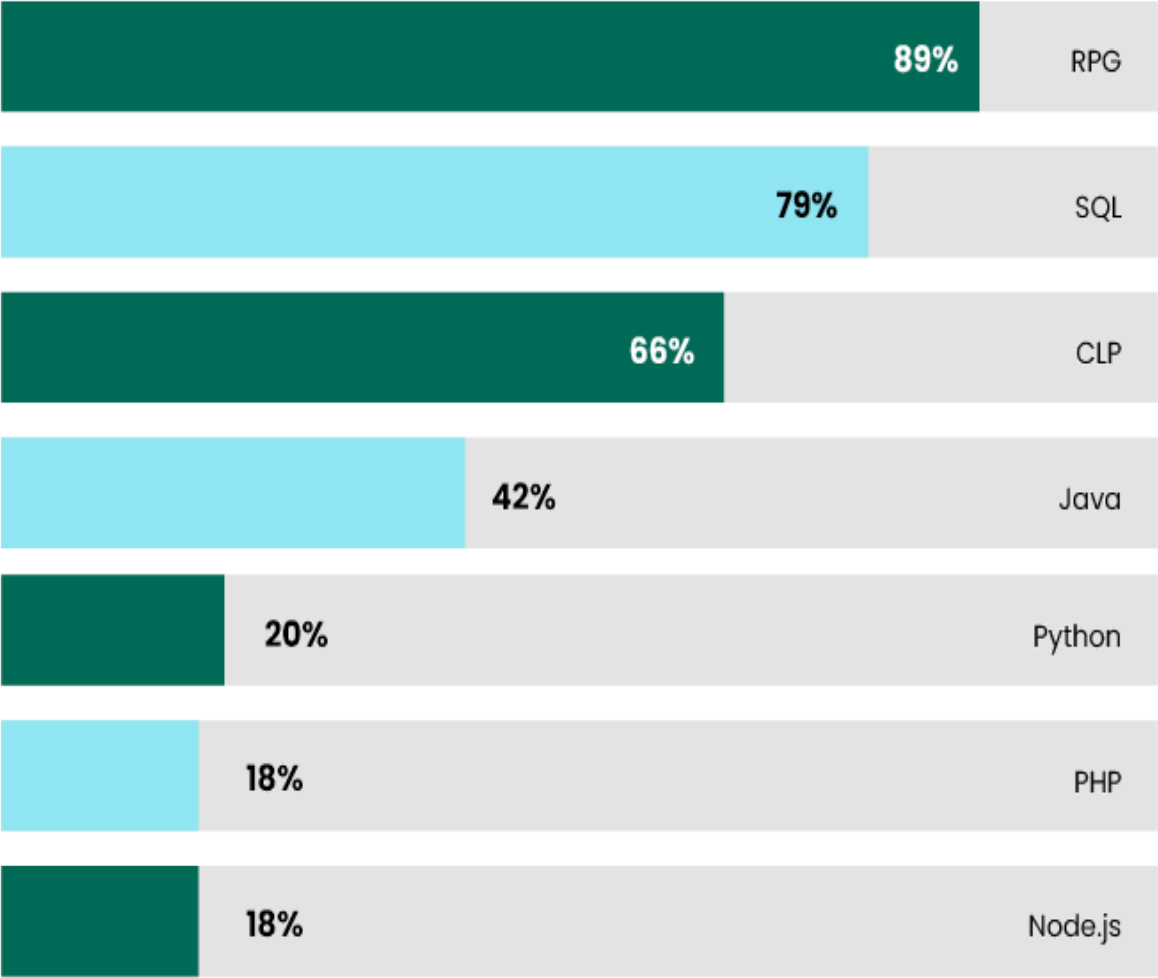
- Help developer write code
- Understand code

What about watsonx Code Assistant for Z?

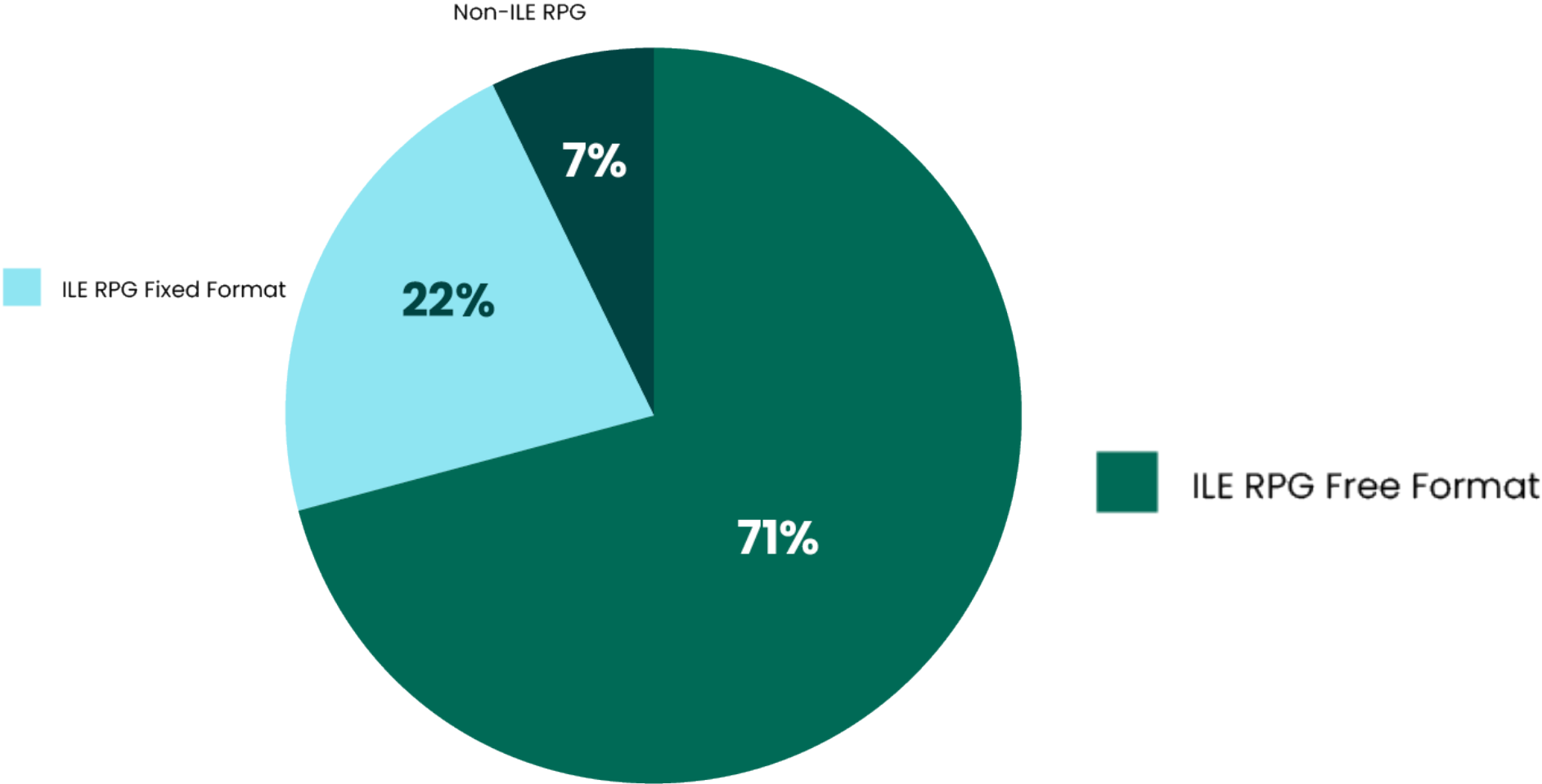
It starts with RPG ... most of the time

Development Languages

Which development languages do you use today for new development on IBM i?



What is your primary version of RPG for new development?



What is generative AI?

Generative AI, sometimes called *gen AI*, is **artificial intelligence** (AI) that can create original content—such as text, images, video, audio or software code—in response to a user’s prompt or request.

IBM i Community: Code Assist Base Features

An IBM i Code assist tool should

- Help programmers work with existing RPG
- Generate modern free-format ILE RPG based on a description
- Examine and Explain existing code
- Write test programs for RPG

What are LLMs?

Large language models (LLMs) are a category of foundation models trained on immense amounts of data making them capable of understanding and generating natural language and other types of content to perform a wide range of tasks.

<https://www.ibm.com/topics/large-language-models>

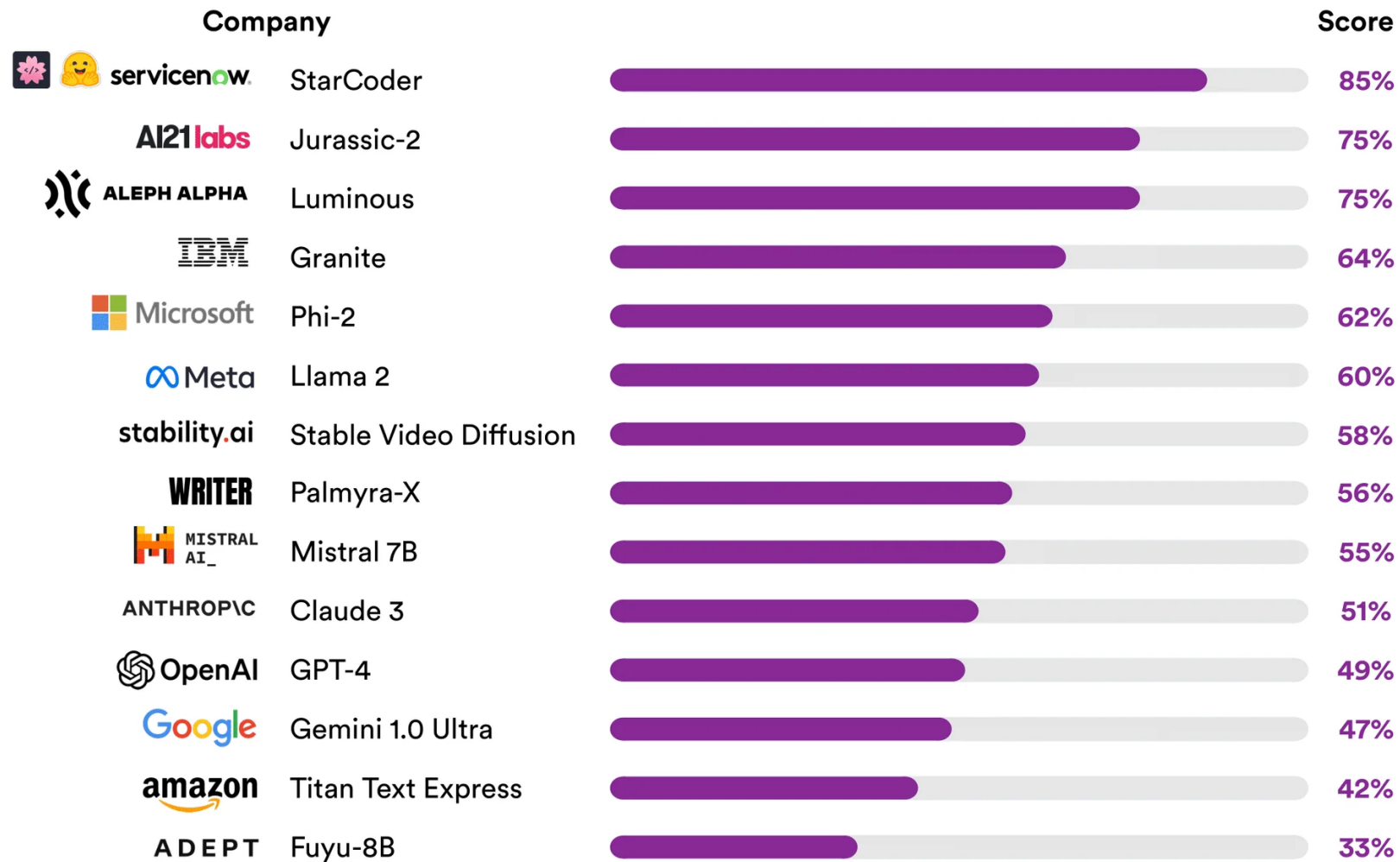
What is a large language model (LLM)?

Large language models (LLMs) are machine learning models that can comprehend and generate human language text. They work by analyzing massive data sets of language.

<https://www.cloudflare.com/learning/ai/what-is-large-language-model/>

Foundation Model Transparency Index Total Scores, May 2024

Source: May 2024 Foundation Model Transparency Index



Foundation Model Transparency Index Scores by Major Dimensions of Transparency, May 2024

Source: May 2024 Foundation Model Transparency Index

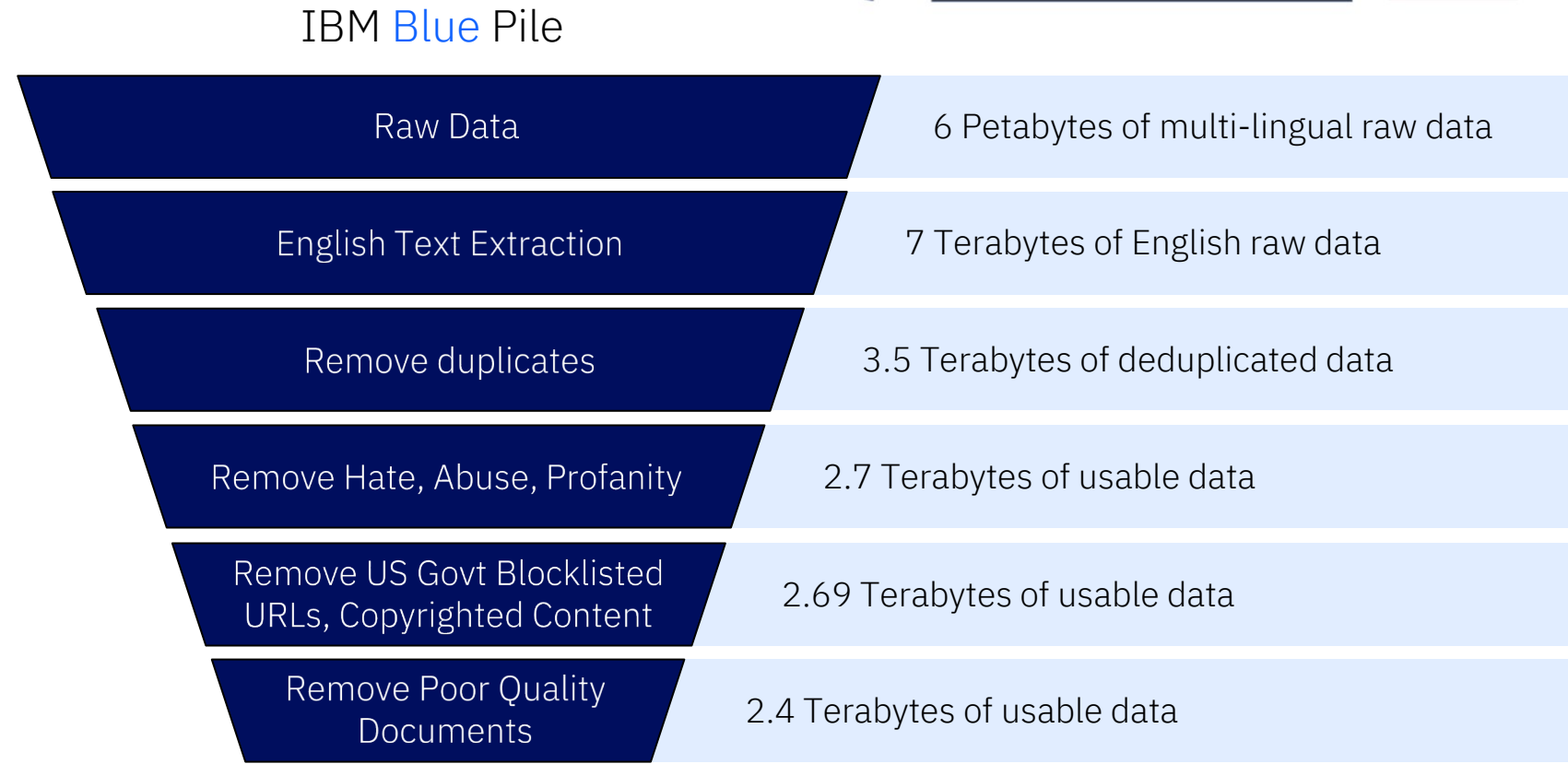
	ADEPT	AI21 labs	ALPHA ALPHA	amazon	ANTHROPIC	servicenow	Google	IBM	Meta	Microsoft	NESTRAL AT	OpenAI	stability.ai	WRITER	Average
	Fuyu-8B	Jurassic-2	Luminous	Titan Text Express	Claude 3	StarCoder	Gemini 1.0 Ultra	Granite	Llama 2	Phi-2	Mistral 7B	GPT-4	Stable Video Diffusion	Palmyra-X	
Data	0%	60%	40%	0%	10%	100%	0%	60%	40%	40%	20%	20%	40%	50%	34%
Labor	0%	43%	71%	14%	14%	100%	29%	43%	29%	100%	100%	14%	100%	43%	50%
Compute	14%	86%	100%	0%	14%	100%	14%	100%	71%	57%	14%	14%	43%	86%	51%
Methods	0%	100%	100%	50%	75%	100%	75%	100%	75%	100%	100%	50%	75%	100%	79%
Model Basics	83%	100%	100%	83%	50%	100%	83%	100%	100%	100%	100%	50%	100%	100%	89%
Model Access	100%	67%	100%	67%	67%	100%	67%	67%	100%	100%	100%	67%	100%	33%	81%
Capabilities	80%	80%	100%	80%	100%	100%	80%	60%	100%	100%	100%	100%	60%	100%	89%
Risks	0%	57%	57%	43%	86%	100%	43%	71%	71%	29%	14%	57%	14%	14%	47%
Mitigations	0%	40%	20%	20%	40%	0%	40%	80%	60%	0%	60%	60%	0%	20%	31%
Distribution	57%	86%	100%	57%	86%	100%	57%	86%	71%	71%	71%	71%	86%	71%	77%
Usage Policy	40%	100%	100%	80%	100%	100%	100%	40%	40%	100%	40%	80%	60%	80%	76%
Feedback	67%	100%	67%	67%	33%	100%	67%	67%	33%	67%	67%	33%	67%	33%	62%
Impact	29%	29%	29%	0%	14%	14%	29%	0%	14%	0%	14%	14%	14%	14%	15%
Average	36%	73%	76%	43%	53%	86%	53%	67%	62%	66%	62%	49%	58%	57%	

IBM Foundation Models



IBM Granite Differentiators

- Smaller, more efficient models without sacrificing accuracy
- Trained on highly governed, cleansed and de-duplicated data
- Provide full auditable data lineage to our clients
- Legal Indemnification for 3rd Party Copyright Claims



Tokenized to 1+ Trillion Tokens for Training



What are Small Language Models?

Small language models are simple and efficient types of [neural networks](#) made for handling language tasks. They work almost as well as bigger models but use far fewer resources and need less computing power.

<https://www.analyticsvidhya.com/blog/2024/05/what-are-small-language-models-slms/>

An IBM i Code assist tool should

- Help programmers work with existing RPG
- Generate modern free-format ILE RPG based on a description
- Examine and Explain existing code
- Write test programs for RPG
- ... and, by the way, transform older RPG into modern, ILE-based free-format

RPG Assistant Concept: Generate RPG

Given the table PAYROLL defined in the SQL below, generate a free form ILE RPG procedure that would calculate the amount of due to a employee with the given employee_id who has worked the given amount of hours. Use the employee's pay rate and tax rate that can be retrieved from the PAYROLL table.

```
CREATE TABLE payroll (  
  employee_id INT NOT NULL,  
  pay_rate DECIMAL(7, 2) NOT NULL,  
  tax_rate DECIMAL(7,2) NOT NULL,  
  PRIMARY KEY (id)
```

Press ⌘+Enter to apply

Generate

Clear Input

Clear Output

RPG Assistant Concept: Generate RPG

Output:

```
1  **free
2
3  dcl-proc calculate_pay export;
4      dcl-pi calculate_pay packed(12:2);
5          employee_id int value;
6          hours packed(5:2) value;
7      end-pi;
8
9      dcl-s pay_rate packed(7:2);
10     dcl-s tax_rate packed(7:2);
11     dcl-s gross_pay packed(12:2);
12     dcl-s net_pay packed(12:2);
13
14     select pay_rate, tax_rate
15     from payroll
16     where employee_id = :employee_id;
17
18     gross_pay = pay_rate * hours;
19     net_pay = gross_pay - (gross_pay * tax_rate);
20
21     return net_pay;
22 end-proc;
```

Code-Text. Instruction: Summarize the next RPG code

```
D wMsgSta 1A
D wMsgStaTxt 10A

C GetMsgStatTxt BEGSR
C SELECT
C WHEN wMsgSta = 'D'
C EVAL wMsgStaTxt = 'Deleted'
C WHEN wMsgSta = 'R'
C EVAL wMsgStaTxt = 'Read'
C WHEN wMsgSta = 'U'
C EVAL wMsgStaTxt = 'Unread'
C ENDSL
C ENDSR
```

The RPGLE code defines a subroutine called "GetMsgStatTxt" that translates a message status code into a textual description. There are two variables declared at the start of the program: 'wMsgSta' which is a one-character variable to hold the message status, and 'wMsgStaTxt' which is a ten-character variable intended to hold the textual description of the status.

The 'GetMsgStatTxt' subroutine uses a 'Select' block to determine what the content of 'wMsgStaTxt' should be based on the value of 'wMsgSta'. If 'wMsgSta' is 'D', 'R', or 'U', the corresponding text ('Deleted', 'Read', or 'Unread', respectively) is assigned to 'wMsgStaTxt'. There are no other actions or alternative cases specified for other potential values of 'wMsgSta'.

RPG Assistant Concept: Generate Unit Test

Given the procedure and DDL below, please write a unit test that will exercise a number of edge cases for this procedure. In the unit include the SQL DML to get the SQL into the desired state.

RPG Assistant Concept: Generate Unit Test

```
Dcl-Proc calculate_pay_test;  
  Dcl-PI calculate_pay_test;  
  End-PI;  
  
  Exec SQL  
    DELETE FROM employee;  
  Exec SQL  
    DELETE FROM payroll;  
  
  Exec SQL  
    INSERT INTO employee (employee_id, pay_rate, tax_rate)  
    VALUES (1, 10.00, 0.20);  
  
  Exec SQL  
    INSERT INTO employee (employee_id, pay_rate, tax_rate)  
    VALUES (2, 15.00, 0.15);  
  
  Exec SQL  
    INSERT INTO employee (employee_id, pay_rate, tax_rate)  
    VALUES (3, 20.00, 0.10);  
  
  Exec SQL  
    INSERT INTO employee (employee_id, pay_rate, tax_rate)  
    VALUES (4, 25.00, 0.05);  
  
  Exec SQL  
    INSERT INTO employee (employee_id, pay_rate, tax_rate)  
    VALUES (5, 30.00, 0.00);
```

RPG Assistant Concept: Generate Unit Test

```
// Test case 1: Employee with ID 1, hours worked = 10  
employee_id = 1;  
hours_worked = 10;  
calculate_pay(employee_id: hours_worked);
```

Exec SQL

```
SELECT pay_date, gross_pay, tax_paid, net_pay  
INTO :pay_date, :gross_pay, :tax_paid, :net_pay  
FROM payroll  
WHERE employee_id = :employee_id;
```

```
AssertEqual(pay_date: %Date(): 'Pay date should be today');  
AssertEqual(gross_pay: 100.00: 'Gross pay should be 100.00');  
AssertEqual(tax_paid: 20.00: 'Tax paid should be 20.00');  
AssertEqual(net_pay: 80.00: 'Net pay should be 80.00');
```

And it keeps going...

How do you train a model to do that? It's not magic!

- Most efficiently, with pairs
 - [Block of Code; Explanation of Block of Code]
 - This sort of pair helps the model learn how to take English (together with other things like SQL) and create code which accomplishes what the English describes.
 - This sort of pair also helps the model look at code and then generate an English description.
 - [Block of Code; Block of Code which Tests the first block]
 - Trains the model to understand how to test various kinds of code.
 - [Block of old code; Block of new code which does the same thing]
 - Trains the model to transform old to new.
- Training based on huge amounts of non-paired code is possible, but far, far more expensive and time-consuming. We will not start with this sort of training, but we will accept code for use later in training.

IBM i Approach – Involve the Community

- IBM i is launching a project to train a “ready for prime time” AI language model-based code assistant
 - Using the RPG code IBM has developed
 - Using code donated by experts & Champions
 - Susan Gantner, Jon Paris, Scott Klement, Jim Buck, Paul Tuohy, Niels Liisberg
 - Yvonne Enselman, Mats Lidström, Koen DeCorte, Steve Bradshaw
 - ... and we hope to have many more!
- Soliciting code for training from the entire IBM i community
 - That could mean YOU!
- The project for a community-trained Large Language Model has begun!

So How Do You Get Involved?

- Let us know you want to help
 - e-mail AlforIBMi@ibm.com
- Agree to the license
 - We'll send it to you – it says we can use your submitted material to train the model
- Submit code
 - Information on how is found at: <https://ibm.github.io/rpg-genai-data/#/>
 - You decide if others (outside IBM) can see & use your code or not
- At some point in the future, we will likely ask for volunteers to actually use the submissions in training & evaluating



Welcome

Overview

How to contribute training data

Training Data Format

Task Description and Examples

Explain Existing RPG

Generate new RPG

Generate Unit Tests for Existing RPG

Fixed to Free Format

How to contribute RPG source

How to make contribution IBM Repo

Training RPG Generative AI

IBM would like to train generative AI models to understand RPG and IBM i application development. We recognize that many customers have valuable assets and expertise and we want to collaborate on building an AI that will help us all.

How to contribute

1. Email AlforIBMi@ibm.com to sign up to contribute to training this AI
2. IBM will reach out and ask you to sign the appropriate legal agreements
3. Once you have signed the appropriate agreements, please provide your github id to the same email AlforIBMi@ibm.com. You can create one [here](#)
4. IBM will add you as a contributor to the git repository <https://github.com/AlforIBMi/rpg-genai-data> where we gathering all of the training data.
5. Now it is time to unleash your expertise. Follow the instructions below
 - We would prefer contributing actual question and answer pairs to train the AI following [these instructions](#). We can start fine tuning our model with this data right away.
 - It will also be useful to contribute RPG source code following [these instructions](#). If we get sufficient source we can do a general unlabeled training of our model.
 - In either case, contributing to the IBM git repository for this data should follow [this process](#)



Welcome

Overview

How to contribute training data

Training Data Format

Task Description and Examples

Explain Existing RPG

Generate new RPG

Generate Unit Tests for Existing RPG

Fixed to Free Format

How to contribute RPG source

How to make contribution IBM Repo

Training RPG Generative AI

IBM would like to train generative AI models to understand RPG and IBM i application development. We recognize that many customers have valuable assets and expertise and we want to collaborate on building an AI that will help us all.

<https://ibm.github.io/rpg-genai-data/#/>

Use cases and examples

The scenarios we are initially focused on are

1. [Explaining RPG in natural language](#)
2. [Prototyping new RPG code from a natural language description](#)
3. [Generating RPG unit tests](#)
4. [Converting Fixed form RPG to free format](#) The latter is not something we expect to use the AI for as there are excellent deterministic tools to do so. But we feel that this is an excellent way to train the AI.
5. Modernize old RPG to new RPG In addition we are very interested in modernizing RPG. While this is not going to be our first deliverable, we are eager to start collecting data to train the AI on this high value transformation. {This is still being defined}

How to contribute

We would prefer contributing actual question and answer pairs to train the AI following [these instructions](#)

It will also be useful to contribute RPG source code following [these instructions](#)

In either case, contributing to the IBM git repository for this data should follow [this process](#)

What have we decided? What's left to decide?

Decided

- Model: We're starting with Granite 8
- Collection of donated training material (RPG code; + "pair" when possible): 3Q 2024
- License requirements: based on open license, but specific to use for training
- Public or Private Contributions – both are allowed

Not Yet Decided

- Timing
- Delivery / Offering
- Pretty much anything else you can think to ask that I haven't mentioned ...

There is so much potential after the first functions!

Future, more advanced versions of the LLM could be trained to address

- Converting S/36 RPG, RPG II, etc. into Free-Format ILE RPG
- Suggesting sections of code where open source languages would help
- Transforming old data definitions and access into modern Db2 & SQL
- Helping modernize when other languages are the starting point
- ... and more!!!

Db2 Data Analytics

- Trend analysis
- Anomaly detection

Operations

- Active monitoring / alerting
- Self-healing

Developer Experience

- Help developer write code
- Understand code

IBM i

continuous innovation
continuous integration

