



Take the power of Copilot on the go with the free mobile app Create images and get tailored answers based on your interests and needs anytime, anywhere

No, thanks Get the Copilot app



Research

Our research

All Microsoft

[Return to Blog Home](#)

Microsoft Research Blog

GraphRAG auto-tuning provides rapid adaptation to new domains

Published September 9, 2024

By [Alonso Guevara Fernández](#), Sr. Software Engineer; Katy Smith, Data Scientist II; [Joshua Bradley](#), Senior Data Scientist; [Darren Edge](#), Senior Director; [Ha Trinh](#), Senior Data Scientist; [Sarah Smith](#), Senior Program Manager; [Ben Cutler](#), Senior Director; [Steven Truitt](#), Principal Program Manager; [Jonathan Larson](#), Senior Principal Data Architect

Share this page



GraphRAG uses large language models (LLMs) to create a comprehensive knowledge graph that details entities and their relationships from any collection of text documents. This graph enables GraphRAG to leverage the semantic structure of the data and generate responses to complex queries that require a broad understanding of the entire text. In previous blog posts, we introduced [GraphRAG](#) and demonstrated how it could be [applied to news articles](#). In this blog post, we show that it can also be tuned to any domain to enhance the quality of the results.

BLOG

[GraphRAG: Unlocking LLM discovery on narrative private data >](#)

BLOG

[GraphRAG: New tool for complex data discovery now on GitHub >](#)

The knowledge graph creation process is called *indexing*. An LLM, guided by a set of domain-specific prompts, reads all the source content and extracts the relevant information, including entities and relationships, which are then used to construct the graph. For example, when analyzing news articles, entities like people, places, and organizations are important. Here, relationship types might include “lives in,” “leads,” and “owns.”

However, each domain has a different set of entity and relationship types. In the field of chemistry, for instance, entity types include molecules, enzymes, and reactions, while relationship types include “catalyzes” and “reduces.” Although our default news domain prompts in GraphRAG can produce a graph when applied to chemistry, they don’t capture the specific content a chemist would expect.

Manually creating and tuning a set of domain-specific prompts is time-consuming. We know, as all the prompts used for news articles were generated manually. To streamline this process, we developed an automated tool that generates domain-specific prompts, which are tuned and ready to use. This tool follows a human-like approach; we provided an LLM with a sample of text data (e.g., 1% of 10,000 chemistry papers) and instructed it to produce the prompts it deemed most applicable to the content. Now, with these automatically generated and tuned prompts, we can immediately apply GraphRAG to a new domain of our choosing, confident that we’ll get high-quality results.

Indexing prompts in GraphRAG

During the indexing process, GraphRAG uses a set of prompts to instruct the LLM as it reads through the source content, extracting and organizing relevant information to construct the knowledge graph. Three of GraphRAG’s main indexing prompts include:

- 1. **Entity and relationship extraction:** Identifies all the entities present and establishes relationships among them.
- 2. **Entity and relationship summarization:** Consolidates instances of entities and their relationships into a single, concise description.
- 3. **Community report generation:** Generates a summary report for each community within the constructed knowledge graph.

These prompts work best when tuned to the domain of the source content. In the rest of this blog post, we focus on domain tuning of the first prompt, “Entity and relationship extraction,” but similar methods apply to the second and third prompts.

Below, Code Sample 1 shows the default few-shot prompt for entity and relationship extraction. This prompt was originally created for news articles and is the default form found in the GraphRAG [GitHub repository](#). The extraction prompt comprises four sections:

- **Extraction instructions:** Provide the LLM with guidance on how to perform extraction.
- **Few-shot examples:** Supply the LLM real examples of the types of entities and relationships worth extracting.
- **Real data:** Serves as a placeholder that is replaced by chunks of source content.
- **Gleanings:** Encourage the LLM, over multiple turns, to extract additional information.

The goal of auto-tuning is to create customized few-shot examples that are appropriate for the given domain. The default prompt, shown in Code Sample 1, provides the LLM with fifteen entity examples and twelve relationship examples, but it is notably restricted to just a few specific entity types: organization, geography, and person. These samples were invented by our team and do not represent real entities.

✓ Code Sample 1: Default prompt

Goal

Given a text document that is potentially relevant to this activity and a list of entity types, identify all entities of those types from the text and all relationships among the identified entities.

Steps

1. Identify all entities. For each identified entity, extract the following information:

- entity_name: Name of the entity, capitalized
- entity_type: One of the following types: [{entity_types}]
- entity_description: Comprehensive description of the entity's attributes and activities

Format each entity as ("entity", , ,)

2. From the entities identified in step 1, identify all pairs of (source_entity, target_entity) that are *clearly related* to each other.

- source_entity: name of the source entity, as identified in step 1
- target_entity: name of the target entity, as identified in step 1
- relationship_description: explanation as to why you think the source entity and the target entity are related to each other
- relationship_strength: a numeric score indicating strength of the relationship between the source entity and target entity

Format each relationship as ("relationship", , , ,)

3. Return output in English as a single list of all the entities and relationships identified in steps 1 and 2. Use {record_delimiter} as the list delimiter.

4. When finished, output {completion_delimiter}

Examples

Example 1:

Entity_types: ORGANIZATION,PERSON

Text:

The Verdantis's Central Institution is scheduled to meet on Monday and Thursday, with the institution planning to release its latest policy decision on Thursday at 1:30 p.m. PDT, followed by a press conference where Central Institution Chair Martin Smith will take questions. Investors expect the Market Strategy Committee to hold its benchmark interest rate steady in a range of 3.5%-3.75%.

Output:

("entity", **CENTRAL INSTITUTION**, ORGANIZATION, The Central Institution is the Federal Reserve of Verdantis, which is setting interest rates on Monday and Thursday)

("entity", **MARTIN SMITH**, PERSON, Martin Smith is the chair of the Central Institution)

("entity", **MARKET STRATEGY COMMITTEE**, ORGANIZATION, The Central Institution committee makes key decisions about interest rates and the growth of Verdantis's money supply)

("relationship", **MARTIN SMITH – CENTRAL INSTITUTION**, Martin Smith is the Chair of the Central Institution and will answer questions at a press conference, 9)

Example 2:

Entity_types: ORGANIZATION

Text:

TechGlobal's (TG) stock skyrocketed in its opening day on the Global Exchange Thursday. But IPO experts warn that the semiconductor corporation's debut on the public markets isn't indicative of how other newly listed companies may perform.

TechGlobal, a formerly public company, was taken private by Vision Holdings in 2014. The well-established chip designer says it powers 85% of premium smartphones.

Output:

("entity", **TECHGLOBAL**, ORGANIZATION, TechGlobal is a stock now listed on the Global Exchange which powers 85% of premium smartphones)

("entity", **VISION HOLDINGS**, ORGANIZATION, Vision Holdings is a firm that previously owned TechGlobal)

("relationship", **TECHGLOBAL – VISION HOLDINGS**, Vision Holdings formerly owned TechGlobal from 2014 until present, 5)

Example 3:

Entity_types: ORGANIZATION,GEO,PERSON

Text:

Five Aurelians jailed for 8 years in Firuzabad and widely regarded as hostages are on their way home to Aurelia.

The swap orchestrated by Quintara was finalized when \$8bn of Firuzi funds were transferred to financial institutions in Krohaara, the capital of Quintara.

The exchange initiated in Firuzabad's capital, Tiruzia, led to the four men and one woman, who are also Firuzi nationals, boarding a chartered flight to Krohaara.

They were welcomed by senior Aurelian officials and are now on their way to Aurelia's capital, Cashion.

The Aurelians include 39-year-old businessman Samuel Namara, who has been held in Tiruzia's Alhamia Prison, as well as journalist Durke Bataglani, 59, and environmentalist Meggie Tazbah, 53, who also holds Bratinas nationality.

Output:

("entity", **FIRUZABAD**, GEO, Firuzabad held Aurelians as hostages)

("entity", **AURELIA**, GEO, Country seeking to release hostages)

("entity", **QUINTARA**, GEO, Country that negotiated a swap of money in exchange for hostages)

("entity", **TIRUZIA**, GEO, Capital of Firuzabad where the Aurelians were being held)

("entity", **KROHAARA**, GEO, Capital city in Quintara)

("entity", **CASHION**, GEO, Capital city in Aurelia)

("entity", **SAMUEL NAMARA**, PERSON, Aurelian who spent time in Tiruzia's Alhamia Prison)

("entity", **ALHAMIA PRISON**, GEO, Prison in Tiruzia)

("entity", **DURKE BATAGLANI**, PERSON, Aurelian journalist who was held hostage)

("entity", **MEGGIE TAZBAH**, PERSON, Bratinas national and environmentalist who was held hostage)

("relationship", **FIRUZABAD – AURELIA**, Firuzabad negotiated a hostage exchange with Aurelia, 2)

("relationship", **QUINTARA – AURELIA**, Quintara brokered the hostage exchange between Firuzabad and Aurelia, 2)

("relationship", **QUINTARA – FIRUZABAD**, Quintara brokered the hostage exchange between Firuzabad and Aurelia, 2)

("relationship", **SAMUEL NAMARA – ALHAMIA PRISON**, Samuel Namara was a prisoner at Alhamia prison, 8)

("relationship", **SAMUEL NAMARA – MEGGIE TAZBAH**, *Samuel Namara and Meggie Tazbah were exchanged in the same hostage release, 2)*
("relationship", **SAMUEL NAMARA – DURKE BATAGLANI**, *Samuel Namara and Durke Bataglani were exchanged in the same hostage release, 2)*
("relationship", **MEGGIE TAZBAH – DURKE BATAGLANI**, *Meggie Tazbah and Durke Bataglani were exchanged in the same hostage release, 2)*
("relationship", **SAMUEL NAMARA – FIRUZABAD**, *Samuel Namara was a hostage in Firuzabad, 2)*
("relationship", **MEGGIE TAZBAH – FIRUZABAD**, *Meggie Tazbah was a hostage in Firuzabad, 2)*
("relationship", **DURKE BATAGLANI – FIRUZABAD**, *Durke Bataglani was a hostage in Firuzabad, 2)*

Real Data
#####

Entity_types: {entity_types}
Text: {input_text}

Output:

Customization can be difficult and time-consuming—in both determining the right set of entities and relationships and in carefully constructing all the prompts for a specific domain. We address these challenges with auto-tuning.

MICROSOFT RESEARCH BLOG



Microsoft Research Forum Episode 3: Globally inclusive and equitable AI, new use cases for AI, and more

In the latest episode of Microsoft Research Forum, researchers explored the importance of globally inclusive and equitable AI, shared updates on AutoGen and MatterGen, presented novel use cases for AI, including industrial applications and the potential of multimodal models to improve assistive technologies.

[Read more >](#)

Auto-tuning architecture

Auto-tuning takes source content and produces an automatically generated set of domain-specific prompts. Figure 1 shows the architecture of the auto-tuning process for our three main indexing prompts.

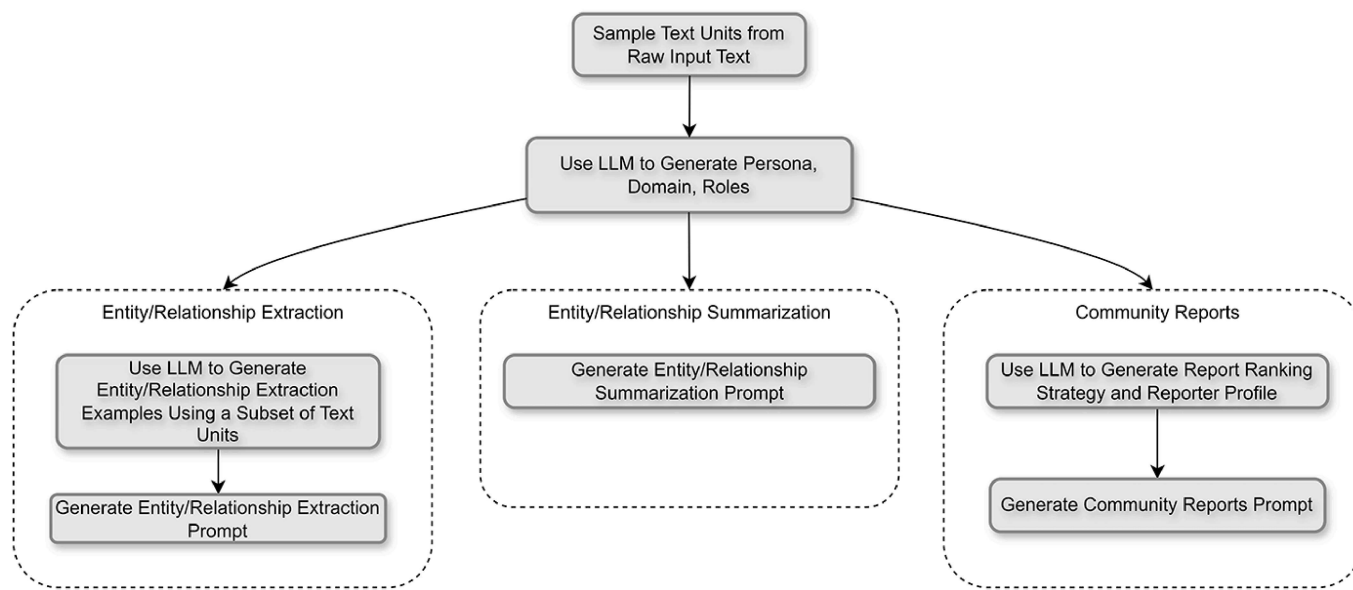


Figure 1. Diagram of the algorithm

We start by sending a sample of the source content to the LLM, which first identifies the domain and then creates an appropriate persona—used with downstream agents to tune the extraction process. Once the domain and persona are established, several processes occur in parallel to create our custom indexing prompts. This way, the few-shot prompts are generated based on the actual domain data and from the persona’s perspective.

To illustrate how this works in practice for entity and relationship extraction, let’s shift to a new domain, the *Behind the Tech* podcast.

Auto-tuning the *Behind the Tech* podcast

Kevin Scott, CTO of Microsoft, hosts a podcast series called *Behind the Tech* where he interviews a wide variety of tech innovators. Given its focus on society and technology, this dataset would benefit from its own set of indexing prompts distinct from general news. While the default prompt works with podcast transcripts, we can achieve much higher precision with customized podcast-tuned prompts.

To demonstrate this, we use Code Sample 2, which contains a sample raw text input chunk from the podcast.

Code Sample 2: Podcast data sample

KEVIN SCOTT: Our guest today is **Ashley Llorens**. Ashley is a scientist, engineer, and hip-hop artist. He worked for two decades at **Johns Hopkins Applied Physics Laboratory**, developing novel AI technologies and served as the founding chief of the lab’s **Intelligent Systems Center**.

He was recently nominated by the **White House Office of Science and Technology Policy** to serve as an AI expert for the **Global Partnership on AI**. Besides his career in engineering, Ashley actually began his career as a hip-hop artist and serves as a voting member of the **Recording Academy** for the **Grammy Awards**.

About a month ago, Ashley joined **Microsoft** as a vice president, distinguished scientist, and managing director for **Microsoft Research**. Welcome to the show, Ashley – and to Microsoft.

ASHLEY LLORENS: Thanks so much, Kevin, great to be here.

The first step in adapting GraphRAG to the target domain is to generate a persona for the LLM to assume when generating examples for each prompt. As it adapts to the domain from the podcast text sample input, the LLM produces the following:

“You are an expert in social network analysis with a focus on technology and innovation communities. You are skilled at mapping and interpreting complex networks, identifying key influencers, and understanding the dynamics of community interactions. You are adept at helping organizations and researchers identify the relations and structure within specific domains, particularly in rapidly evolving fields like technology and innovation.”

Using the persona as part of the prompt, along with the text sample input, we allow the LLM to generate the entity and relationship-extraction prompt, including custom examples. Our indexing prompt is now automatically tuned to our new domain, as shown in Code Sample 3.

> Code Sample 3: Auto-tuned prompt

Here, the automatically generated prompt using the sample content from Code Sample 2 identifies fourteen entity examples across six different entity types (person, location, group, concept, field, and geography) and eight relationship examples.

To assess how this impacts the extraction of the entire dataset, we used both the default and the auto-tuned prompt to generate the entity and relationship outputs. Before we explain the results, let's review the default prompt's outputs, which produced seven entities and six relationships, as shown in Code Sample 4.

Code Sample 4: Default extraction output

Entities and Relationships:

("entity", **ASHLEY LLORENS**, PERSON, *Ashley Llorens is a scientist, engineer, hip-hop artist, and vice president at Microsoft. He has worked at Johns Hopkins Applied Physics Laboratory and was nominated by the White House to serve as an AI expert for the Global Partnership on AI. He is also a voting member of the Recording Academy for the Grammy Awards.*)

("entity", **JOHNS HOPKINS APPLIED PHYSICS LABORATORY**, ORGANIZATION, *An institution where Ashley Llorens worked for two decades developing novel AI technologies and served as the founding chief of the intelligent systems center.*)

("entity", **WHITE HOUSE OFFICE OF SCIENCE AND TECHNOLOGY POLICY**, ORGANIZATION, *A U.S. government office that nominated Ashley Llorens to serve as an AI expert for the Global Partnership on AI.*)

("entity", **GLOBAL PARTNERSHIP ON AI**, ORGANIZATION, *An international initiative focused on AI where Ashley Llorens serves as an expert.*)

("entity", **RECORDING ACADEMY**, ORGANIZATION, *An organization responsible for the Grammy Awards, of which Ashley Llorens is a voting member.*)

("entity", **MICROSOFT**, ORGANIZATION, *A multinational technology company where Ashley Llorens joined as a vice president, distinguished scientist, and managing director for Microsoft Research.*)

("entity", **KEVIN SCOTT**, PERSON, *The host of the podcast where Ashley Llorens was interviewed and welcomed to Microsoft.*)

("relationship", **ASHLEY LLORENS – JOHNS HOPKINS APPLIED PHYSICS LABORATORY**, *Ashley Llorens worked at Johns Hopkins Applied Physics Laboratory for two decades, 9)*

("relationship", **ASHLEY LLORENS – WHITE HOUSE OFFICE OF SCIENCE AND TECHNOLOGY POLICY**, *Ashley Llorens was nominated by the White House Office of Science and Technology Policy to serve as an AI expert, 8)*

("relationship", **ASHLEY LLORENS – GLOBAL PARTNERSHIP ON AI**, *Ashley Llorens serves as an AI expert for the Global Partnership on AI, 8)*

("relationship", **ASHLEY LLORENS – RECORDING ACADEMY**, *Ashley Llorens is a voting member of the Recording Academy for the Grammy Awards, 7)*

("relationship", **ASHLEY LLORENS – MICROSOFT**, *Ashley Llorens joined Microsoft as a vice president and managing director for Microsoft Research, 9)*

("relationship", **KEVIN SCOTT – ASHLEY LLORENS**, *Kevin Scott interviewed Ashley Llorens on the podcast and welcomed him to Microsoft, 8)*

Using the auto-tuned, domain-specific, automatically generated prompt, we achieved a deeper extraction, producing nine entities and eight relationships, as shown below in Code Sample 5.

Code Sample 5: Auto-tuned extraction output

Entities and Relationships:

("entity", **ASHLEY LLORENS**, PERSON, *Ashley Llorens is a scientist, engineer, hip-hop artist, and executive at Microsoft. He has a background in AI technology development and has been involved in the music industry as a hip-hop artist and Grammy voting member*)

("entity", **KEVIN SCOTT**, PERSON, *Kevin Scott is the host of the podcast where Ashley Llorens is a guest. He engages with Ashley about his background and career*)

("entity", **JOHNS HOPKINS APPLIED PHYSICS LABORATORY**, ORGANIZATION, *An institution where Ashley Llorens worked for two decades, developing novel AI technologies and leading the intelligent systems center*)

("entity", **WHITE HOUSE OFFICE OF SCIENCE AND TECHNOLOGY POLICY**, ORGANIZATION, *A governmental office that nominated Ashley Llorens to serve as an AI expert for the Global Partnership on AI*)

("entity", **GLOBAL PARTNERSHIP ON AI**, ORGANIZATION, *An international initiative focused on AI where Ashley Llorens serves as an expert.*)

("entity", **RECORDING ACADEMY**, ORGANIZATION, *An organization responsible for the Grammy Awards, of which Ashley Llorens is a voting member.*)

("entity", **MICROSOFT**, ORGANIZATION, A major technology company where Ashley Llorens recently joined as a vice president, distinguished scientist, and managing director for Microsoft Research)

("entity", **CHICAGO**, LOCATION, The city where Ashley Llorens grew up, specifically mentioned as the south side and south suburbs, which influenced his interest in music and technology)

("entity", **HIP-HOP**, MUSIC GENRE, A music genre that significantly influenced Ashley Llorens during his childhood in Chicago, leading him to pursue a career in music alongside his technical career)

("relationship", **ASHLEY LLORENS – JOHNS HOPKINS APPLIED PHYSICS LABORATORY**, Ashley Llorens worked at Johns Hopkins Applied Physics Laboratory for two decades, developing AI technologies, 9)

("relationship", **ASHLEY LLORENS – WHITE HOUSE OFFICE OF SCIENCE AND TECHNOLOGY POLICY**, Ashley Llorens was nominated by the White House Office of Science and Technology Policy to serve as an AI expert, 9)

("relationship", **ASHLEY LLORENS – GLOBAL PARTNERSHIP ON AI**, Ashley Llorens serves as an AI expert for the Global Partnership on AI, 9)

("relationship", **ASHLEY LLORENS – RECORDING ACADEMY**, Ashley Llorens is a voting member of the Recording Academy for the Grammy Awards, 7)

("relationship", **ASHLEY LLORENS – MICROSOFT**, Ashley Llorens recently joined Microsoft as a vice president and managing director for Microsoft Research, 9)

("relationship", **ASHLEY LLORENS – CHICAGO**, Ashley Llorens grew up in Chicago, which influenced his early interest in music, particularly hip-hop, 7)

("relationship", **ASHLEY LLORENS – HIP-HOP**, Ashley Llorens was deeply influenced by hip-hop music during his upbringing in Chicago, leading him to pursue a career in music, 8)

("relationship", **KEVIN SCOTT – ASHLEY LLORENS**, Kevin Scott hosts Ashley Llorens on the podcast, discussing his background and career transitions, 7)

Compared with the default prompt, the auto-tuned prompt is an improvement, with more entities and more relationships, providing a more comprehensive view of our data. One key difference between this output and the output from the default prompt is the expansion in entity types being extracted. The default prompt is limited to three example types: organization, geography, and person. However, the auto-tuned prompt expands to more example types derived from the sample input text: organization, person, location, and music genre.

Putting it all together

We can observe a clear difference in the final outputs after using these auto-tuned prompts for indexing the podcast source data. To measure this difference, we compared the size of the resulting knowledge graphs using default with auto-tuned prompts. The following results were achieved while keeping all parameters constant between both runs and using GPT4-Turbo:

	Entities	Relationships	Communities
Default prompt	1796	2851	352
Auto-tuned prompt	4896	8210	1027

As shown, the use of auto-tuning yields a significantly larger knowledge graph. For example, a prompt that looks for molecules will extract much more from a chemistry dataset than one that looks for people and places. More communities in the knowledge graph means that it can better serve global search queries. While the size of the knowledge graph can be a proxy for its utility, the true measure of utility is observed in end-to-end queries. To demonstrate this, we provide the following side-by-side comparison of a question using the default and auto-tuned prompts.

. . .

We asked GraphRAG the following question: "What’s the relationship between Ashley Llorens and Chicago?"

First, we got the following results using the knowledge graph built with the default prompt. We then used the knowledge graph built with our auto-tuned prompt:

Default query results

"I am sorry but I am unable to answer this question given the provided data."

Auto-tuned query results

****Origin and Influence****: Ashley Llorens has a significant connection to Chicago, which has deeply influenced his personal and professional life. Growing up in a city known for its vibrant music scene, particularly hip-hop, has shaped his musical pursuits and artistic expression."

*“**Professional Impact**”: His background and experiences in Chicago have carried over into his professional environment, notably at the Applied Physics Laboratory, where he has integrated elements of hip-hop culture [Data: Reports (940)].”*

With auto-tuned indexing prompts, our knowledge graph became more representative of the dataset’s entities and relationships, enabling it to yield a valid response to the query.

Advancing auto-tuning and expanding GraphRag’s capabilities

Looking ahead, we’re exploring other methods to build on this auto-tuning work. We’re excited to investigate new approaches for creating the core GraphRAG knowledge graph and are also studying ways to measure and evaluate the quality of these graph structures. Additionally, we’re researching methods to better assess performance so that we can identify the types of queries where GraphRAG provides unique value. This includes evaluating human-generated versus auto-tuned prompts, as well as exploring potential improvements to the auto-tuner.

Overall, these new auto-tuner developments make GraphRAG much more accessible and turnkey. We hope this auto-tuning work removes many of the challenges involved when working with new datasets. We invite you to try out these capabilities yourself using GraphRAG’s [core library](#) and our Azure-based solution accelerator, available on [GitHub](#).

[Try out GraphRAG](#)

[Try out GraphRAG Accelerator](#)

Related publications

[From Local to Global: A Graph RAG Approach to Query-Focused Summarization](#) >

Meet the authors



Alonso Guevara Fernández

Sr. Software Engineer

[Learn more](#) >



Katy Smith

Data Scientist II