hr | avatar
Illuminating Talent

Test Results and Interview Guide

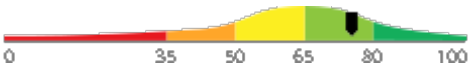
Candidate: **Elizabeth Wantsajob**
Assessment: Natural Language Processing (NLP) Concepts
Completed: July 2, 2026
Prepared for: Sara Maple
Example Company

What's Included

- Overall Score
- Competency Summary Table
- Comparison Matrix
- Detailed Competency Results with Interview Guide

Important Note: The Natural Language Processing (NLP) Concepts assessment measures key factors related to high performance and tenure in this job. Attribute types measured vary by test, but can include cognitive ability, skills, knowledge, personality characteristics, emotional intelligence, and past behavioral history. This report includes a one page summary, followed by detailed results with an embedded interview guide. Note that these results should always be used as a part of a balanced candidate selection process that includes independent evaluation steps, such as interviews and reference checks.

Overall

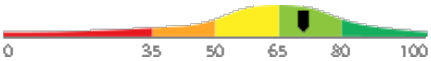
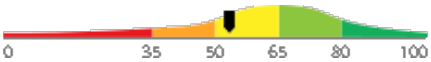
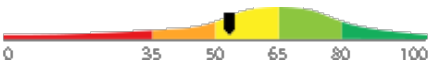
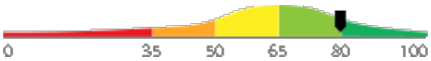
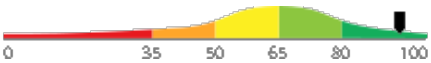
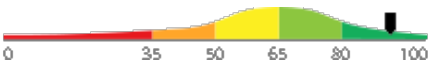
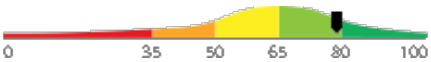
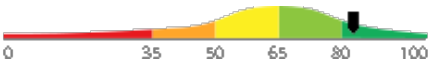
Candidate	Score	Interpretation
Elizabeth Wantsajob	75	

beth.wantsajob@gmail.com
Natural Language Processing (NLP) Concepts
July 2, 2026

The candidate demonstrates a solid and well-rounded understanding of NLP underlying concepts and design principles, including text preprocessing, feature representation methods, pipeline construction, classification techniques, and model evaluation. This level of competency is consistent with a capable entry-level to mid-level professional who can contribute meaningfully to NLP projects with minimal supervision, though continued development in advanced topics may further strengthen performance.

Key
- Candidate Score
- Higher Risk
- Lower Risk

Competency Summary

Competency	Score	Interpretation
Skills/Knowledge (relates to immediate readiness)		
Model Evaluation Metrics	71	
Text Preprocessing (Free Text Responses)	53	
Text Representation (Free Text Responses)	53	
NLP Pipelines	80	
Named Entity Recognition (NER)	94	
Text Classification and Sentiment Analysis	91	
Text Preprocessing	78	
Text Representation	82	

Comparison

Percentile scores indicate how the candidate compares to other test-takers within various groups. The candidate scored equal to or better than the fraction of test-takers indicated by the percentile.

Test-Taker Group	Percentile	0 10 20 30 40 50 60 70 80 90 100
Global	75th	
North America	62nd	
United States	62nd	
Example Company	69th	

Artificial Intelligence (AI) Generated Scores

This table includes one or more scores derived from a large language model AI query. AI-derived scores are non-deterministic. That is, they are not precisely repeatable. Therefore, these scores should always be treated as supplementary information and should never be used exclusively or compared to hard cutoff values.

Estimated Value	Score	Confidence	Interpretation
Knowledge, Skills, and Abilities Summary	-	-	Summary Points (AI): • (Generic Text for Sample Report) Strong performer in Drag and Drop Files tasks, indicating comfort with file management and basic computer interactions. • Demonstrates solid numerical accuracy in Recognizing and Confirming Numbers, a valuable asset in detail-oriented roles. • Moderate overall performance in Analytical Thinking and Attention to Detail, with adequate grammar skills but room for improvement. • Struggles with Reading and Analyzing Problems, which may limit effectiveness in roles requiring critical reading and complex problem-solving. • Lowest performance in Navigating Between Screens, suggesting difficulty with multi-screen software workflows that could impact productivity in computer-intensive roles. Narrative (AI): Elizabeth Wantsajob demonstrates a mixed profile of knowledge, skills, and abilities across the assessed competencies. Elizabeth shows a strong aptitude in Drag and Drop Files, performing well on this technical task and suggesting she is comfortable with this type of computer interaction. This is a notable strength that would translate well into roles requiring file management and basic computer navigation tasks. In the area of Analytical Thinking and Attention to Detail, Elizabeth performs at a moderate level. She demonstrates solid ability in Recognizing and Confirming Numbers, which suggests she is careful and accurate when working with numerical data — a valuable skill in detail-oriented work environments. Her Grammar performance is adequate but leaves room for improvement, indicating she may occasionally make written communication errors. Her weakest area within this competency is Reading and Analyzing Problems, where she struggled to consistently interpret and work through written problem scenarios. This may impact her effectiveness in roles that require critical reading, written comprehension, or complex problem-solving. Elizabeth's most significant area for development is Navigating Between Screens, where she scored considerably lower than the other competencies. This suggests she may have difficulty efficiently moving through software interfaces or multi-screen workflows, which could slow productivity in roles that rely heavily on navigating computer applications or data entry systems. Overall, Elizabeth brings some useful technical strengths, particularly in file management and numerical accuracy, but would benefit from targeted development in software navigation and analytical problem-solving to be fully effective in roles that demand these skills. Computed on: April 2, 2026, 11:09:49PM EDT

Detail

Candidate:	Elizabeth Wantsajob, beth.wantsajob@gmail.com
Assessment:	Natural Language Processing (NLP) Concepts
Authorized:	July 2, 2026, by Sara Maple, Example Company, qamailsaram.mike@hravatar.com
Started:	July 2, 2026, 5:22:59PM EDT
Completed:	July 2, 2026, 5:22:59PM EDT
Overall Score:	75

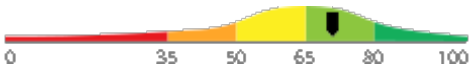
Knowledge and Skills Detail

This section contains a list of job-related knowledge areas and skills that have been evaluated. Low scores in these areas often indicate that additional learning may be required before top performance can be achieved.

Detail | Interview Guide

Model Evaluation Metrics

Score: 71



Description:

Evaluating NLP models requires using the right metrics to measure how well a model performs on a given task. Precision, recall, and F1 score are the most commonly used metrics, each capturing a different aspect of model accuracy and error.

Interpretation:

Candidate should achieve above average job performance in this area with little or no training.

The candidate exhibits a solid understanding of NLP model evaluation metrics, including the practical distinctions and trade-offs between precision, recall, and F1 score. They are likely capable of selecting and applying appropriate metrics for a variety of NLP tasks with moderate independence.

What is the F1 score, and in what situation would it be a more useful metric than accuracy alone?

1	2	3	4	5
Cannot define the F1 score or describes it as simply the average of precision and recall without further explanation.		Correctly defines F1 as the harmonic mean of precision and recall but gives only a basic explanation of when to use it.		Defines F1 score accurately, explains why it is preferred over accuracy for imbalanced datasets, and gives a concrete illustrative example.

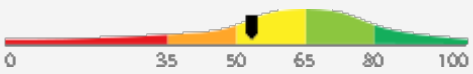
Can you explain what precision and recall measure, and why both are important when evaluating an NLP model?

1	2	3	4	5
Cannot define precision or recall, or significantly confuses the two concepts.		Correctly defines precision and recall individually but cannot explain why both are needed together or how they relate to each other.		Clearly defines both metrics, explains the trade-off between them, and describes a scenario where one might be prioritized over the other.

Detail

Text Preprocessing (Free Text Responses)

Score: 53



Description:

Covers the end-to-end process of planning, building, testing, and deploying AI-enabled applications for both internal staff and external customers. Includes managing iteration cycles, versioning, model monitoring, and coordinating cross-functional teams through each phase of the product lifecycle.

Interpretation:

The candidate exhibits average writing skills, which can hinder high performance in some jobs.

The candidate possesses a moderate understanding of AI product management, demonstrating basic familiarity with lifecycle management, strategic assessment, and process orchestration, though proficiency across these areas is inconsistent. With targeted coaching and hands-on experience, this individual has the potential to develop into a capable contributor in managing AI-enabled application initiatives.

Overall AI Score:	60.0
High words per minute detected while composing one or more essays:	27.3 words per minute. Possible copy/paste or use of AI tools. Average WPM while composing is about 15.
AI Confidence Level:	80
Argument Strength (AI):	70.0
Clarity and Coherence (AI):	80.0
Match with Ideal Response (AI):	60.0
Other Errors per 100 Words:	0.0
Spelling errors per 100 words:	0.0

Please see below to view the essay submitted.

Interview Guide

Describe a time you managed or contributed to an AI product through multiple lifecycle stages. What were the most significant challenges you encountered between phases, and how did you address them?

1	2	3	4	5
Candidate provides a generic or superficial example that lacks detail about AI-specific lifecycle challenges. Does not clearly articulate their personal role or the decisions they made between phases.		Candidate shares a relevant example with reasonable detail, identifying at least one meaningful challenge such as stakeholder alignment or testing delays. However, the response may lack specificity about how AI-related factors (e.g., model performance, data readiness) influenced lifecycle decisions.		Candidate provides a detailed, concrete example that demonstrates ownership across multiple lifecycle phases. Clearly describes AI-specific challenges such as model validation failures, shifting requirements, or deployment infrastructure issues, and articulates the specific actions they took to resolve them and keep the product on track.

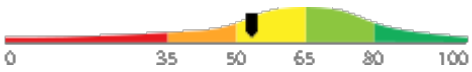
Can you walk me through the basic stages you would follow to take an AI-enabled product from an initial idea to a live deployment?

1	2	3	4	5
Candidate provides a vague or incomplete description of the lifecycle, omitting key phases such as testing, validation, or deployment. May conflate AI product development with general software development without acknowledging AI-specific considerations like model training or data pipelines.		Candidate identifies the major phases (discovery, development, testing, deployment) and acknowledges some AI-specific considerations, but struggles to articulate how the phases connect or how cross-functional teams are coordinated throughout.		Candidate clearly outlines a structured lifecycle including discovery, requirements, development, model validation, testing, deployment, and monitoring. Demonstrates awareness of AI-specific challenges such as data quality, model drift, and iterative retraining, and explains how they would coordinate stakeholders across phases.

Detail

Text Representation (Free Text Responses)

Score: 53



Description:

Covers the end-to-end process of planning, building, testing, and deploying AI-enabled applications for both internal staff and external customers. Includes managing iteration cycles, versioning, model monitoring, and coordinating cross-functional teams through each phase of the product lifecycle.

Interpretation:

The candidate exhibits average writing skills, which can hinder high performance in some jobs.

The candidate possesses a moderate understanding of AI product management, demonstrating basic familiarity with lifecycle management, strategic assessment, and process orchestration, though proficiency across these areas is inconsistent. With targeted coaching and hands-on experience, this individual has the potential to develop into a capable contributor in managing AI-enabled application initiatives.

Overall AI Score:	60.0
High words per minute detected while composing one or more essays:	27.3 words per minute. Possible copy/paste or use of AI tools. Average WPM while composing is about 15.
AI Confidence Level:	80
Argument Strength (AI):	70.0
Clarity and Coherence (AI):	80.0
Match with Ideal Response (AI):	60.0
Other Errors per 100 Words:	0.0
Spelling errors per 100 words:	0.0

Please see below to view the essay submitted.

Interview Guide

Describe a time you managed or contributed to an AI product through multiple lifecycle stages. What were the most significant challenges you encountered between phases, and how did you address them?

1	2	3	4	5
Candidate provides a generic or superficial example that lacks detail about AI-specific lifecycle challenges. Does not clearly articulate their personal role or the decisions they made between phases.		Candidate shares a relevant example with reasonable detail, identifying at least one meaningful challenge such as stakeholder alignment or testing delays. However, the response may lack specificity about how AI-related factors (e.g., model performance, data readiness) influenced lifecycle decisions.		Candidate provides a detailed, concrete example that demonstrates ownership across multiple lifecycle phases. Clearly describes AI-specific challenges such as model validation failures, shifting requirements, or deployment infrastructure issues, and articulates the specific actions they took to resolve them and keep the product on track.

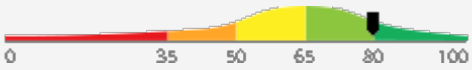
Can you walk me through the basic stages you would follow to take an AI-enabled product from an initial idea to a live deployment?

1	2	3	4	5
Candidate provides a vague or incomplete description of the lifecycle, omitting key phases such as testing, validation, or deployment. May conflate AI product development with general software development without acknowledging AI-specific considerations like model training or data pipelines.		Candidate identifies the major phases (discovery, development, testing, deployment) and acknowledges some AI-specific considerations, but struggles to articulate how the phases connect or how cross-functional teams are coordinated throughout.		Candidate clearly outlines a structured lifecycle including discovery, requirements, development, model validation, testing, deployment, and monitoring. Demonstrates awareness of AI-specific challenges such as data quality, model drift, and iterative retraining, and explains how they would coordinate stakeholders across phases.

Detail

NLP Pipelines

Score: 80



Description:

An NLP pipeline is a sequence of connected processing steps that transform raw text into a useful output. Understanding how components such as preprocessing, feature extraction, and model inference are ordered and connected is essential for building and maintaining NLP systems.

Interpretation:

Candidate should achieve superior job performance in this area with little or no training.

The candidate demonstrates a comprehensive and advanced understanding of NLP pipelines, reflecting a strong command of how all major components are sequenced, connected, and optimized within end-to-end NLP systems. This level of proficiency indicates the ability to independently design, implement, and maintain sophisticated NLP pipeline architectures.

Interview Guide

What challenges might arise when connecting components in an NLP pipeline, and how would you approach diagnosing a problem in one?

1	2	3	4	5
Cannot identify common pipeline issues or gives only a vague, generic answer about debugging.		Identifies one or two issues such as mismatched data formats but does not describe a systematic approach to diagnosing problems.		Identifies multiple potential issues, describes a step-by-step diagnostic approach, and mentions checking outputs at each stage.

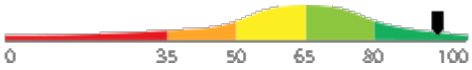
Can you walk me through what an NLP pipeline is and describe the basic steps it might include?

1	2	3	4	5
Cannot describe a pipeline or lists steps in an illogical order without explanation.		Describes a pipeline as a series of steps and names some correct components but does not explain how they connect or why the order matters.		Clearly describes a pipeline, explains why steps must occur in a specific order, and names key components from input to output.

Detail

Named Entity Recognition (NER)

Score: 94



Description:

Named entity recognition (NER) is the task of identifying and classifying specific items in text, such as names of people, organizations, locations, and dates. It is widely used to extract structured information from unstructured text.

Interpretation:

Candidate should achieve superior job performance in this area with little or no training.

The candidate demonstrates an advanced and comprehensive mastery of Named Entity Recognition concepts and their application within the broader field of Natural Language Processing. They are highly capable of identifying, classifying, and extracting named entities from complex, unstructured text and can be expected to contribute at an expert level in this domain.

Interview Guide

What factors can make named entity recognition difficult, and how might the domain or context of text affect NER performance?

1	2	3	4	5
Cannot identify any meaningful challenges or does not connect domain context to NER performance.		Identifies one or two general challenges but does not clearly explain how domain-specific language affects recognition accuracy.		Identifies multiple challenges such as ambiguous names and domain jargon, and explains how models trained on one domain may underperform in another.

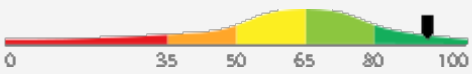
Can you explain what named entity recognition is and describe the types of entities it is typically used to identify?

1	2	3	4	5
Cannot define NER or describes it inaccurately without reference to identifying specific named items in text.		Correctly defines NER and names one or two entity types but cannot explain how it is applied in practice.		Clearly defines NER, lists multiple entity types such as person, organization, and location, and describes a practical use case.

Detail

Text Classification and Sentiment Analysis

Score: 91



Description:

Text classification is the task of assigning predefined labels or categories to text documents, such as spam detection or topic labeling. Sentiment analysis is a specific type of classification that determines whether text expresses a positive, negative, or neutral opinion.

Interpretation:

Candidate should achieve superior job performance in this area with little or no training.

The candidate demonstrates an advanced and comprehensive mastery of text classification and sentiment analysis within NLP. They show a deep understanding of classification methodologies, predefined label assignment, and the nuanced detection of positive, negative, and neutral sentiment, positioning them as a highly competent practitioner in this domain.

Interview Guide

How does sentiment analysis work, and what are some of the challenges that can make it difficult to accurately determine the sentiment of a piece of text?

1	2	3	4	5
Cannot explain how sentiment analysis works or cannot identify any meaningful challenges.		Explains the basic goal of sentiment analysis but identifies only obvious challenges such as slang without deeper insight.		Explains the mechanism of sentiment analysis and identifies nuanced challenges such as sarcasm, negation, and domain-specific language.

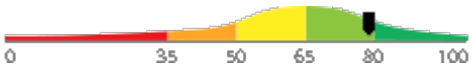
Can you explain what text classification is and give an example of how it might be used in a real application?

1	2	3	4	5
Cannot define text classification or provides an example that does not involve categorizing text.		Correctly defines text classification and gives a basic example but does not explain how the model assigns labels.		Clearly defines text classification, gives a relevant real-world example, and briefly explains how labels are assigned based on learned patterns.

Detail

Text Preprocessing

Score: 78



Description:

Text preprocessing involves preparing raw text for use in NLP models. This includes techniques such as tokenization (splitting text into words or sentences), stopword removal, stemming, and lemmatization to reduce words to a standard form.

Interpretation:

Candidate should achieve above average job performance in this area with little or no training.

The candidate exhibits a solid and well-rounded understanding of Natural Language Processing principles, including text preprocessing, representation methods, classification techniques, sentiment analysis, and model evaluation. Minor gaps in advanced or specialized areas may exist, but the candidate is well-positioned to contribute effectively to NLP-related tasks with limited guidance.

Interview Guide

What is the difference between stemming and lemmatization, and when would you choose one over the other?

1	2	3	4	5
Cannot distinguish between stemming and lemmatization or confuses the two concepts entirely.		Correctly distinguishes the two but cannot clearly explain when to prefer one approach over the other.		Clearly explains the difference, notes lemmatization produces valid words while stemming is faster, and gives a practical use-case comparison.

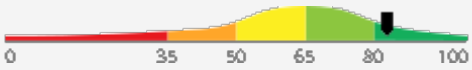
Can you describe what tokenization is and why it is used as a first step when working with text data?

1	2	3	4	5
Cannot define tokenization or gives a vague, inaccurate description with no practical context.		Correctly defines tokenization and mentions it splits text into units, but does not explain why it matters.		Clearly defines tokenization, explains its role in preparing text for models, and may contrast word vs. sentence tokenization.

Detail	Interview Guide

Text Representation

Score: 82



Description:

Text representation refers to the methods used to convert text into numerical formats that machine learning models can process. Common approaches include bag-of-words, TF-IDF, and word embeddings, each capturing different aspects of meaning and context.

Interpretation:

Candidate should achieve superior job performance in this area with little or no training.

The candidate exhibits an advanced and comprehensive understanding of text representation within NLP, demonstrating strong command of how various techniques convert text into numerical formats and the trade-offs between them. They are well-equipped to independently apply, evaluate, and optimize text representation strategies in complex machine learning workflows.

How does TF-IDF improve on a simple word count representation, and what problem is it designed to address?

1	2	3	4	5
Cannot explain TF-IDF or confuses it with basic word frequency counting.		Explains that TF-IDF reduces the weight of common words but cannot fully articulate how the IDF component works.		Clearly explains both term frequency and inverse document frequency components and how together they highlight distinctive words in a document.

Can you explain what a bag-of-words model is and how it represents a piece of text?

1	2	3	4	5
Cannot describe bag-of-words or provides a significantly incorrect explanation.		Correctly describes bag-of-words as word counts but does not mention its limitations, such as ignoring word order.		Accurately explains bag-of-words, notes it ignores word order and context, and contrasts it with more advanced representations.

Free Text Responses

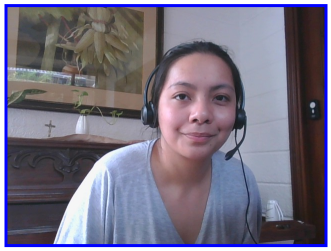
During the assessment, the candidate was asked to answer one or more questions using text, audio, video, or an uploaded text file. Their responses are included below for review.

Question or Task	Response
After an AI product is deployed, what is model monitoring and why is it a necessary part of the product lifecycle?	Model monitoring is a technique for ensuring that the model does not wander or become overtrained after an extended period of repeated queries that have the same or similar prompts. This is very important for preventing hallucinization. It's also a key aspect of any guardrails strategy. **Comments (AI):** The answer is clear and coherent but lacks depth in explaining the importance of model monitoring. The phrase 'hallucinization' is not commonly used in this context and may confuse readers. The answer could be improved by providing more specific examples of model performance metrics and how they are tracked. The argument strength is moderate as it does not fully explain why model monitoring is necessary in the product lifecycle. **Misspelled Words:** guardrails (1), hallucinization (1)

Identity Confirmation Photos

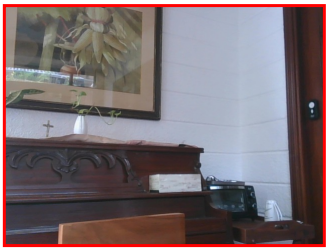
The following photos of the candidate and any identification were uploaded during the assessment session.

Photo Analysis Results	
- Risk:	Medium risk of cheating based on image inconsistencies
- Percent match among processed faces	100%
- Total images processed	17
- Total images with valid faces	14 (82%)
- Total pairs of faces compared	13
- Pairs in which faces matched	13 (100%)

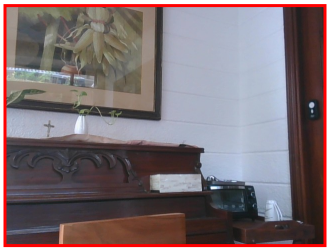


Pre/Post-Test Photo

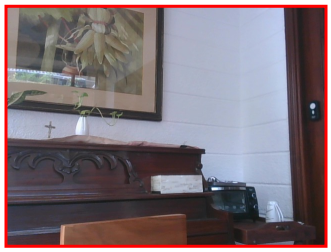
ID Photo



In-Test Error Detected (No Face Detected)

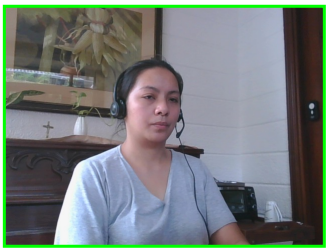


In-Test Error Detected (No Face Detected)

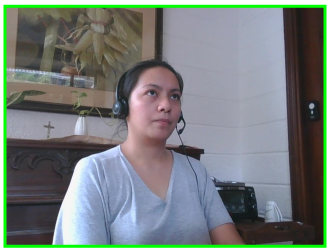


In-Test Error Detected (No Face Detected)

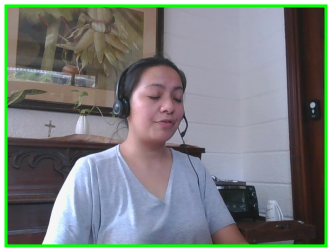
In-Test Photo



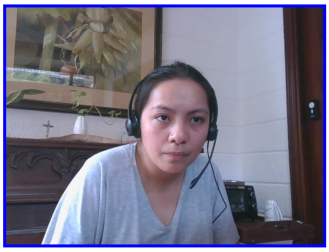
In-Test Photo



In-Test Photo



In-Test Photo



Pre/Post-Test Photo

Resume or CV

Summary Updated on

Motivated career professional with extensive experience in office administration and management. Proven track record of improving efficiency, reducing costs, and enhancing office operations through strategic initiatives and technology implementation.

Objective

I am seeking a role where I can use my many skills and my exceptional judgment and empathy for customers to make a difference to a growing company.

Education

- Associate of Applied Science in Office Administration, Portland Community College, 2020

Experience

- General Office Clerk, Paramount Office Management, 09/2023 – Present
- Administrative Assistant, Global Enterprises Inc., 04/2021 – 08/2023
- Administrative Assistant, Innovative Business Solutions Ltd., 07/2019 – 03/2021

Other Qualifications

- Microsoft Office Specialist (MOS) Certification
- Certified Administrative Professional (CAP)
- International Association of Administrative Professionals (IAAP) Certification

Report Preparation Notes

- Hiring decisions should never be based on a single source of information. The most effective use of this assessment report is as a part of a multi-faceted program of candidate evaluation that includes resume review, interviews, and reference checks.
- Overall vs Percentiles Scores: The overall score reflects the success in the test, based on the mean (average) and standard deviation of the test scores. The percentile score reflects the percentage of test-takers who scored equal or below this overall score. We recommend you use the Overall Score as your primary evaluation criteria. However, percentile scores can often be useful in comparing specific candidates against one another and with a group, such as for test takers in a certain organization or within a certain account.
- Note that comparison information is calculated based on completed instances of this assessment at that time the assessment is scored. As additional instances are completed, the comparative data may change. You can always update a report to the current values by clicking on 'Recalculate Percentiles' within the online results viewing pages at www.hravatar.com.
- Most competency scores are norm-based, which means that they can be interpreted in terms of their distance from the average or mean score. For all scales, a score equal to the mean receives a score of 65 and scores above and below this value are set so that a score change of 15 equals one standard deviation.
- For linear competencies, higher is better across the entire scale. For these scales a score between 65 and 80 (light green) represents 0 to 1 standard deviation above the mean and a score above 80 (dark green) represents more than one standard deviation above the mean. Similarly, a score of 50 - 65 (yellow) represents 0 to 1 standard deviation below the mean, while a score of 35 - 50 (orange) equates to 1 to 2 standard deviations below the mean, and a score below 35 represents more than 2 standard deviations below the mean.
- Sim ID: 20891-1, Key: 0-0, Rpt: 68, Prd: 9711, Created: 2026-07-02 17:23 EDT
- UA: Mozilla/5.0 (Windows NT 6.3; Trident/7.0; Touch; rv:11.0) like Gecko

Score Calculation Detail

The following table provides a summary of how the overall score was calculated from each of the individual competency scores. First, all competency scores are calculated on a scale of 0-100. Note that some competencies use their color category rather than their actual numeric score in the overall calculation. For these, a standard score associated with the assigned color category is used in the overall score calculation rather than the actual numeric score. This is reflected in the ''Score Value Used'' column. Next, a weighted average of scores is computed using individual competency weights, typically set using job analysis data provided by the US Government Occupational Information Network (O*Net).

Competency	Score	How applied to overall	Score Value Used	Weight (%)
Model Evaluation Metrics	71.1634	Numeric Score	71.1634	12.5000
NLP Pipelines	80.2342	Numeric Score	80.2342	12.5000
Named Entity Recognition (NER)	94.0627	Numeric Score	94.0627	12.5000
Text Classification and Sentiment Analysis	91.9043	Numeric Score	91.9043	12.5000
Text Preprocessing	78.9937	Numeric Score	78.9937	12.5000
Text Preprocessing (Free Text Responses)	53.8624	Numeric Score	53.8624	12.5000
Text Representation	82.9260	Numeric Score	82.9260	12.5000
Text Representation (Free Text Responses)	53.8624	Numeric Score	53.8624	12.5000
Weighted Average:				75.8761
Final Overall Score:				75

Notes

(This area is intentionally blank - it's reserved as space for your notes.)