PRODIGY
INTELLIGENCE

Prodigy intelligence White Paper on an Evidence-Based Framework for AI-Assisted Deception-Risk Analysis

Executive Summary

Coyote is an AI-assisted behavioral intelligence and deception-risk analysis platform developed by Prodigy Intelligence. It analyzes text, audio, and video to identify linguistic, behavioral, and para-verbal deception-risk indicators. Its purpose is not to declare whether a person is lying. Instead, Coyote provides probability-based assessments that help users identify statements, themes, or portions of a communication that may warrant further examination.

Internal controlled testing of Coyote's evaluated text-analysis model has reported classification accuracy of up to 90.9% under the specific conditions and datasets used for that testing. This result provides evidence of criterion-related validity but should not be interpreted as a universal accuracy rate. Performance may vary according to language, statement length, subject matter, recording quality, transcription accuracy, population, communication context, and differences between controlled and operational environments.

Coyote should therefore be understood as a decision-support system whose validity rests on three elements: alignment with established deception research, measurable model performance, and responsible human interpretation.

The Scientific Basis for Coyote

Deception is not associated with a single universal behavior or "tell." Research examining verbal and nonverbal indicators has consistently found that individual cues tend to be weak, context-dependent, and unreliable when interpreted in isolation. There is no scientifically established equivalent of a behavioral "Pinocchio's nose."

This limitation is reflected in human performance. A major meta-analysis found that people identify truths and lies with an average accuracy of approximately 54%, only moderately above chance. Human evaluators are also vulnerable to confirmation bias, cultural assumptions, overconfidence, fatigue, and reliance on stereotypical behaviors such as gaze avoidance or nervous movement.

Coyote addresses these limitations by evaluating multiple indicators systematically rather than relying on one behavior. Depending on the input, the platform can examine

language patterns, changes in narrative structure, distancing language, response characteristics, and relationships between statements. Its outputs are intended to focus an analyst's attention, not replace investigative reasoning or establish deception as fact.

Machine-learning research supports the potential value of this approach. A 2023 systematic review identified 81 studies applying machine learning to deception detection, with reported performance ranging from approximately 51% to 100%. The review noted that instances where accuracy of the machine learning model was 100% the model likely suffered from training data overfit, and was unlikely to obtain this level of accuracy on data from outside the specific dataset utilized to training the model. The review also found considerable variation in datasets, modalities, experimental conditions, and evaluation procedures. These findings demonstrate both the promise of computational methods and the importance of avoiding broad generalizations from a single performance result.

Studies have also shown that verbal cues of deception to be more diagnostic than para-verbal and behavioral signals of deception. As a result the of this the currently Coyote models have limited increased in accuracy from the text model to the audio and video models. This is likely due to the level of accuracy achieved by the current text model as well as the higher level of diagnostic capability of linguistic cues of deception.

AI Model Training and Development

Coyote was developed using theory-informed transformer models, meaning that its artificial intelligence models were not trained solely to discover statistical correlations within data. The models were designed and trained with reference to established theory and research concerning deception analysis. This approach helped guide the selection, organization, and interpretation of relevant features while allowing the transformer architecture to identify complex relationships that may not be apparent through traditional rule-based analysis.

For text-based analysis, transformer models were trained to evaluate communications within context rather than treating individual words or phrases as isolated indicators. The models examine relationships among sentences, changes in language, narrative structure, internal consistency, evasiveness, distancing language, certainty, and other potentially relevant features. No individual word, phrase, or behavioral characteristic is treated as proof of deception. Instead, the models evaluate combinations of indicators and the context in which they occur to produce a probability-based assessment.

Coyote's audio-analysis model incorporated Wav2Vec, a neural-network architecture developed for learning representations directly from speech. Wav2Vec enables the model to extract meaningful characteristics from audio recordings while

reducing dependence on manually defined acoustic features. These representations can be used to examine speech related patterns and can be combined with the linguistic information contained in the speaker's transcribed statements. The audio model therefore considers relevant information from both the spoken content and the characteristics of the recorded speech, subject to recording quality and the limitations of the available data.

During development, model performance was evaluated using labelled examples in which the applicable truth or deception category had been established for the purposes of training or testing. The data were divided so that model development and performance evaluation could be conducted on separate examples, reducing the risk that reported results reflected memorization of the training material. Model outputs were then compared with the established labels to measure classification performance and identify areas in which further training, calibration, or refinement was required.

The theory-informed training approach was intended to improve scientific coherence, interpretability, and generalizability. However, theoretical alignment does not itself guarantee that every prediction will be correct. Model performance depends on the quality, representativeness, and labelling of the training data, as well as the language, population, context, recording conditions, and type of communication being analyzed. Coyote's outputs are therefore presented as decision-support assessments that require human interpretation and should be considered alongside contextual information and independent evidence.

Evidence of Validity

Coyote's validity can be considered across several complementary dimensions.

Construct validity concerns whether the platform measures features meaningfully related to the concept being assessed. Coyote is construct-aligned because it examines linguistic, para-verbal, and behavioral characteristics identified in deception research. It does not assume that any single feature proves deception. Instead, indicators are combined into a probabilistic assessment.

Criterion-related validity concerns how closely the model's classifications correspond with independently established truthful or deceptive examples. Controlled internal testing of the evaluated Coyote text model reported accuracy of up to 90.3%, and the Coyote audio model reported accuracy of up to 90.9%. This indicates that, within the evaluated dataset and testing conditions, the model successfully differentiated the labelled classes at a substantially greater rate than chance.

Operational validity concerns whether the system performs reliably in its intended environment. Coyote supports operational use by producing structured reports, identifying the locations of elevated indicators, presenting probability-based risk levels, and preserving the underlying material for human review. These features enable analysts to evaluate the basis of an assessment rather than receiving an unexplained binary conclusion.

Limitations and Generalizability

Coyote's reported performance should not be assumed to apply equally to every person, language, industry, or communication setting. Deception behavior can vary according to culture, language proficiency, stress, neurodiversity, communication style, motivation, preparation, and the consequences associated with the interaction. The currently available Coyote models were only training on truthful and deceptive instances in English, and Coyote should not be utilized for deception analysis in languages other than English. The currently Coyote models were training on a variety of English speakers to include English as a second language, accent variation, communication style, motivation, preparation, and the consequences associated with truthful and deceptive instances.

Very short statements may provide insufficient linguistic information for a stable assessment, and internal testing has shown that Coyote's accuracy rate decreases significantly if there is less than 5 words to analyze in a text analysis. Audio and video analysis may also be affected by background noise, overlapping speech, speaker-identification errors, compression, transcription quality, camera placement, or missing contextual information.

For these reasons, Coyote does not determine truthfulness, guilt, intent, fraud, misconduct, or criminal responsibility. Its results should not be used as the sole basis for employment, legal, financial, immigration, disciplinary, medical, law-enforcement, or other high-impact decisions.

Responsible Validation and Use

The National Institute of Standards and Technology (NIST) states that trustworthy AI requires ongoing testing and monitoring to confirm that a system continues to perform as intended. NIST also recommends documenting the conditions under which a system has been validated and clearly identifying limitations on generalizability.

Consistent with these principles, Coyote's continuing validation includes:

- Segregated training, validation, and unseen test datasets;
- Testing by modality, language, industry, population, and use case;

- Comparison with human and computational baselines;
- Error, bias, calibration, and model-drift analysis; and
- Version-specific documentation of datasets, methods, metrics, and limitations.

Error, bias, calibration, and model-drift analysis have been conducted on all Coyote models, and public facing models have been selected for their accuracy, and where applicable have been training to identify and counter bias. Though not all biases have been accounted for the biases that Prodigy intelligence has identified as the primary causes of analytical drift have been addressed. Version-specific documentation of datasets are maintained as proprietary information as providing the identification of the composition of these datasets may allow other organizations to recreate Coyote.

C**onclusion**

The available evidence supports Coyote as a scientifically informed and technically promising deception-risk decision-support platform. Internal testing reporting accuracy of up to 90.9% provides preliminary evidence that the evaluated models can identify meaningful differences between labelled truthful and deceptive communications under controlled conditions.

The scientifically appropriate conclusion is not that Coyote can determine whether someone is lying. Rather, Coyote can systematically identify patterns associated with elevated deception risk, prioritize material for further review, and provide professionals with structured information that may improve the consistency and efficiency of human analysis.

Coyote's strongest and most defensible application is therefore human-in-the-loop behavioral intelligence, combining computational pattern recognition with context, corroborating evidence, professional judgment, and responsible oversight.

For further information please contact Contact@ProdigyIntel.com or visit www.ProdigyIntel.com

References

DePaulo, B., Lindsay J., Malone B., Muhlenbruck L, Charlton K, Cooper H. (2003). Cues to deception. Psychol Bull. 129(1), 74-118. doi: 10.1037/0033-2909.129.1.74. PMID: 12555795.

Global Deception Research Team. (2006). A world of lies. *Journal of cross-cultural psychology*, *37*(1), 60-74.

Hartwig M, Bond CF. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. Psychol Bull. 137(4), 643-659. doi: 10.1037/a0023589.

Hauch V, Blandón-Gitlin I, Masip J, Sporer SL. (2014). Are computers effective lie detectors? A meta-analysis of linguistic cues to deception. Pers Soc Psychol Rev. 19(4), 307-342. doi: 10.1177/1088868314556539.

Hyde, S., Bachura, E., Bundy, J., Gretz, R., Sanders, G. (2023). The tangled webs we weave: Examining the effects of CEO deception on analyst recommendations. doi: 10.1002/smj.3546

Mann S, Vrij A, Bull R. (2004) Detecting true lies: police officers' ability to detect suspects' lies. J Appl Psychol. 89(1), 137-49. doi: 10.1037/0021-9010.89.1.137.

National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework 1.0. https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10

Sporer, S.L. and Schwandt, B. (2006). Paraverbal indicators of deception: a meta-analytic synthesis. Appl. Cognit. Psychol., 20: 421-446. doi: 10.1002/acp.1190

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons.

Vrij, A., Hartwig, M., Granhag, P. (2019). Reading Lies: Nonverbal Communication and Deception. *Annual Review of Psychology* 70:295-317. doi: 10.1146/annurev-psych-010418-103135