DECEPTIcON: Bridging Gaps in In-the-Wild Deception Research

Berat Biçer, Bahadır Durmaz, Serhat Aras, and Hamdi Dibeklioğlu, *Member, IEEE*

Abstract—We present *DECEPTIcON: Deception on Iconic Individuals* dataset; a novel, multimodal, in-the-wild compilation of data from 100 distinct individuals who are influential figures from contemporary American history. Our goals in collecting *DECEPTIcON* are to address data sparsity and annotation bias commonly seen in existing datasets by ensuring consistent visual quality across samples, while delegating annotations to fact-checks for objectivity and preventing human bias. We also provide the results of the first-ever baselines for *DECEPTIcON*, to be used as reference by future studies. With *DECEPTIcON*, we aspire to contribute not only to the understanding of deception dynamics but also to the enhancement of public discourse and democratic ideals through insightful analysis.

Index Terms—political deception, fact-checking, deceit detection, affective computing, multimodal data analysis

1 Introduction

Deception is a frequently-observed phenomena in our daily lives yet it's nature is not well understood by the common people. This is mainly due to misconceptions around how there are tell-tale sins of deceptive intent based on body language [1], [2]. While this is partially true, according to criminal psychology, no singular behavioral cue is sufficient to ascertain deception [3], [4]. Furthermore, most studies conducted on deception analyse deception in laboratory settings, which makes it difficult to invoke deceptive intent in participants. This makes studying high-stakes scenarios (situations where due to strong emotional activation behavioral leaks of deception are clearly visible) [5] challenging. On the other hand, in-the-wild studies, which try and capture deception from outside the laboratory environment, have a litany of issues in regards to data consistency and quality; which may be undesirable from an analysis point of view.

In this paper, we introduce a novel dataset **DECEPTIcON: Deception on Iconic Individuals**, which seeks to address three main problems commonly seen in the literature: (1) Due to the challenging nature of in-the-wild data collection, datasets collected this way tend to be small in quantity, which limits their applicability in deep-learning based multimodal or unimodal behavioral analysis. (2) Due to a lack of universal data collection setup, in-the-wild datasets exhibit significant inter-sample variations in quality due to occlusion, subject's apparent movement, etc. (3) Deceptive intent is annotated by researchers either through a collective agreement or by behavioral scientists. This is a labor-intensive and error-prone process that is also very slow, limiting how much data can be collected.

In *DECEPTIcON*, we collected multimodal (visual, audio, and speech) data samples from 100 distinct individuals from contemporary American history (mainly between 2016 and 2021). These individuals exhibit a broader media presence compared to any random person (senators, elected officials, etc.). Our goals include conducting in-the-wild analysis of deceptive intent based on these subjects, reigniting interest in the field with a novel, high-quality, multimodal database annotated through fact-checks to reduce human bias in deception intent annotations. Furthermore, our analytical focus extends beyond mere detection, as we aspire to analyse the trustworthiness of persons of public interest and the veracity of their statements. This endeavor serves a broader purpose — ensuring accountability for the benefit of the public and the integrity of modern democratic systems. With *DECEPTIcON*, we aspire to contribute not only to the understanding of deception dynamics but also to the enhancement of public discourse and democratic ideals through insightful analysis.

Contributions. This paper introduces the *DECEPTIcON* dataset, a novel multimodal compilation of data from 100 distinct individuals who are influential figures from contemporary American history. With *DECEPTIcON*, we aim to address issues of data sparsity and annotation bias commonly found in existing datasets. We achieve this by ensuring consistent visual quality and a high quantity of samples, while delegating annotations to fact-checks for objectivity. Additionally, we provide the results of baselines for *DECEPTIcON*, serving as a reference for future studies. Our analysis focuses on studying low-stakes deception in the wild, with a particular emphasis on political deception, contributing to the betterment of society as a whole by promoting accountability.

Organization. The paper is organized as follows: After a brief background on political deception and fact-checking in Section 2, we present a literature survey in Section 3. Then, we discuss dataset collection process in Section 4. In Section

- *B. Biçer is with the Department of Computer Engineering, Bilkent University, Ankara, Türkiye (email: berat.bicer@bilkent.edu.tr).*
- *B. Durmaz is with N26, Berlin working as a Software Engineer (email: bdurmazse@gmail.com).*
- *S. Aras is with the Department of Computer Engineering, Başkent University, Ankara, Türkiye and is a Senior Software Engineer at OBSS, Türkiye (email: srharas@gmail.com).*
- *H. Dibeklioğlu is with the Department of Computer Engineering, Bilkent University, Ankara, Türkiye (e-mail: dibeklioglu@cs.bilkent.edu.tr).*

Manuscript received XX XX, XXX; revised XX XX, XXXX.