Review of a Doctoral Thesis at FIT BUT

Doctoral thesis (hereinafter referred to as "thesis"), title of the thesis:

Artificial Intelligence in Computer Security

Name of the doctoral student (hereinafter referred to as "student"), name and surname:

Ing. Anton Firc

Name and institution of the reviewer (full name of the reviewer, full name and country of the institution):

Professor Nicholas Evans
EURECOM
Campus SophiaTech
450 Route des Chappes
06410 Biot
FRANCE

I. Thesis

Appropriateness and relevance

The work addresses the now-widely acknowledged threats of speech deepfakes. The thesis outlines the security issues surrounding computer security, the use of deepfakes in social engineering attacks, fake news, fraud and identity theft, etc. The problem of deepfake detection or, since the thesis covers more than just detection, deepfake defence, is both appropriate and relevant as a focus for doctoral research as society, industry, and governments, especially within the EU, are currently struggling to respond to the misuse of synthetic media. The work is relevant to speech technology research, which is the examiner's area of expertise, but the novel contributions are related to applied computing and cybersecurity. The thesis also cites the relevance of the research to national priorities in the Czech Republic.

A summary of the contributions of the thesis

The research objectives were to provide a structured cybersecurity assessment of the risks posed by speech deepfakes and hence to bridge the claimed gap in existing research that has focused predominantly on detection algorithms without sufficient consideration of attacker models, human factors, and broader protection strategies. Specific goals are outlined in the introduction chapter and then repeated in the conclusions chapter, followed by a detailed account of the related contributions. These highlight the low cost and high accessibility of attacks, the exaggerated ability of human listeners to identify deepfakes, a framework for the reliable comparison and evaluation of competing detection solutions, and an extended protection model. It would hence appear that the stated goals have been achieved.

The student's own contributions to the work, or rather to specific, related publications, are detailed in Section 3.6. The stated contributions are in the order of 50% for seven of the student's eight publications that are most relevant to the thesis work. The student is first author for five of these publications and second author for the remaining three. The examiner notes the absence of [A6] in the list as presented in Section 3.6, as well as in Table 3.4, even if it is included in the similar list presented in the bibliography. A

list of other publications, while also related to deepfake detection, is included in Section 3.7. Presumably, the listed papers, to some of which the student also had a strong contribution and is also first author, are less relevant to the specific thesis contributions.

Novelty and significance:

While deepfake detection is already a heavily researched area, the student's work is innovative in shifting the focus from purely algorithmic development towards a comprehensive cybersecurity perspective. With a background in speech technology, the examiner appreciates the treatment of cybersecurity concepts and approaches in Section 2.3, but would have appreciated a more detailed discussion of how standards, threat and attacker modelling, and red teaming and security testing came to shape the thesis work. Without it, the novel research contributions, struggle somewhat to stand out later in Section 3. In this respect, the examiner was expecting more substantive significance, or differences to the approaches reported in the speech-related literature. Precisely how the research activities and threat or risk analysis presented in Section 3.1 build upon what was already known previously, particularly from the perspective of the speech community rather than the cybersecurity perspective, is not as clear as it could be.

The examiner missed the motivation to use an image-based detector for speech deepfake detection work presented initially in Section 3.2, especially when there would have been numerous speech-specific solutions available as open-source when this work was performed. I disagree with some (though certainly not all) of the claims regarding the shortcomings of previous databases and evaluation frameworks, especially since it is not entirely clear how the reported comparison framework overcomes them. Even so, the evaluation framework and use of augmentations, as illustrated in Figure 3.6, has merit and is also an original contribution. The examiner also appreciates the detailed comparisons of different detection solutions from which there are some interesting findings, some consistent with previously reported results, others less so. The findings show some shortcomings in the current evaluation paradigms with which I agree.

The findings presented in Section 3.3 are at clear odds with previously reported results and is the first work, to the best of the examiner's knowledge, to show the influence of prior sensitisation to deepfakes in the ability of a human lister to detect attacks. The implications are significant. Though it is reassuring to observe the robustness of previous detection solutions to the latest speech synthesis techniques, the contributions of the work presented in Section 3.5 appears to be more conceptual rather than technical.

Evaluation of the formal aspects of the thesis:

The thesis is well-structured and meets the expectations for a PhD thesis. As stated in Section 1.2, the thesis is a commented collection of eight peer-reviewed publications. The examiner appreciates the summary style and the concise presentation of the student's contributions in Chapter 3. The included publications are properly integrated and clearly referenced, and the candidate's contribution to each is stated explicitly. From the perspective of academic writing, the thesis is of a satisfactory language standard, though with occasional, somewhat casual style, and some minor repetition in the structure. The presentation of metrics in Chapter 2 suggests a lack of familiarity with some of the more recent metrics used for the evaluation of deepfake detection solutions, and especially those used for the evaluation of automatic speaker verification solutions. The examiner also finds the title of the thesis to be far too generic and vague, bearing no reference to speech or audio, generative speech technologies, or to deepfake risks,

threats or defences. The examiner nonetheless acknowledges his unfamiliarity with convention in the Czech Republic, which might account for the use of such a non-descript title.

Quality of publications

The core results of the thesis have been published in internationally recognised, high-quality venues. The student has authored or co-authored eight peer-reviewed publications, including two Q1 journal articles indexed in Scopus/SJR, one Q2 journal article, and several CORE-ranked conference papers including one rank A, in addition to other publications in B-ranked or equivalent venues. Based upon his own background and experience as a PhD examiner, a number of publications at speech conferences such as Interspeech or ICASSP (the venues for a substantial number of cited works) would normally be expected for a thesis in this area of research, though he accepts that the claimed contributions relate to cybersecurity, and not to the speech community. In any case, by international standards, the student's publication record meets the requirements for the award of a PhD.

II. Student's overall achievements

Overall R&D activities evaluation:

The thesis and associated publications demonstrate that the candidate has developed satisfactory scientific knowledge and creative ability. He has formulated research questions, designed novel frameworks for attacker modelling and deepfake detection evaluation, and produced original empirical findings. His publication record in Q1/Q2 journals and CORE-ranked conferences confirms both research competence and the capacity to link theory with practical applied computing and cybersecurity.

III. Conclusion

In the examiner's opinion, the thesis *Artificial Intelligence in Computer Security* by Ing. Anton Firc, together with the student's publication record, meets the generally accepted requirements for doctoral-level research. For this reason, and in light of additional remarks above, the examiner therefore recommends that the student be authorised to defend his work in view of obtaining his PhD degree.

Biot, France, 3rd Sept. 2025

Signature of the reviewer: