



Report on the Pilot of the Netflix Classification Tool

August 2018



Disclaimer

The material in this report is of a general nature and should not be regarded as legal advice or relied on for assistance in any particular circumstance or emergency situation. In any important matter, you should seek appropriate independent professional advice in relation to your own circumstances. The Commonwealth accepts no responsibility or liability for any damage, loss or expense incurred as a result of the reliance on information contained in this report.

This report has been prepared for consultation purposes only and does not indicate the Commonwealth's commitment to a particular course of action. Additionally, any third party views or recommendations included in this report do not reflect the views of the Commonwealth, or indicate its commitment to a particular course of action.

Copyright

© Commonwealth of Australia 2018



The material in this report is licensed under a Creative Commons Attribution—4.0 International license, with the exception of:

- the Commonwealth Coat of Arms
- this Department's logo
- any third party material
- any material protected by a trademark, and
- any images and/or photographs.

More information on this CC BY license is set out as follows:

- Creative Commons website—www.creativecommons.org
- Attribution 4.0 international (CC by 4.0)—www.creativecommons.org/licenses/by/4.0.

Enquiries about this license and any use of this discussion paper can be sent to: copyright@communications.gov.au.

Third party copyright

The Department has made all reasonable efforts to clearly identify material where the copyright is owned by a third party. Permission may need to be obtained from third parties to re-use their material.

Attribution

The CC BY licence is a standard form licence agreement that allows you to copy and redistribute the material in any medium or format, as well as remix, transform, and build upon the material, on the condition that you provide a link to the licence, you indicate if changes were made, and you attribute the material as follows:

Licensed from the Commonwealth of Australia under a Creative Commons Attribution 4.0 International licence.

Enquiries about the use of any material in this publication can be sent to: copyright@communications.gov.au.

Using the Commonwealth Coat of Arms

Guidelines for using the Commonwealth Coat of Arms are available from the Department of Prime Minister and Cabinet website at www.pmc.gov.au/government/its-honour.

s 22 - irrelevant to request



Executive summary

In December 2016, the Minister for Communications and the Arts, Senator the Hon Mitch Fifield (the Minister) launched a pilot of the Netflix Classification Tool (the Tool) to produce Australian classification ratings and consumer advice (classification decisions) for films and television series available online in Australia via the Netflix streaming service. The pilot was administered by the Department of Communications and the Arts (the Department) in consultation with key stakeholders - Netflix International B.V (Netflix) and the Classification Board (the Board).

The pilot period was from December 2016 to May 2018. It consisted of two phases, a diagnostic phase and an evaluation phase where classification decisions of the Tool were reviewed by staff assessors and the Board. A number of key performance measures were developed to assess the effectiveness of the Tool to generate classification decisions that are 'broadly consistent' with Australian community standards and classification decisions of the Board.

s 22 - irrelevant to request

1.2. The Netflix Classification Tool

On 11 October 2016, the Minister approved a pilot of the Tool that was designed and developed by Netflix. The pilot commenced in December 2016. Approval for the ongoing use of the Tool would be subject to a satisfactory evaluation of the Tool. To administer the pilot, the Department developed a Memorandum of Understanding (MOU) with Netflix¹, with the pilot to be administered by the Classification Branch (the Branch) of the Department.

The Tool has been designed to produce classification decisions that are in accordance with the Act, the National Classification Code (the Code) and the Guidelines for the Classification of Films (Film Guidelines).

1.3. Pilot of the Netflix Tool

The pilot operated from 4 December 2016 to 30 May 2018. The pilot assessed the overall performance of the Tool in generating classification decisions that are 'broadly consistent' with Australian community standards and with classification decisions made by the Board. The pilot considered a range of performance measures, agreed between the Department and Netflix, which are outlined below.

1.4. Key Performance Measures

The MOU and the Approval Instrument² identify the Key Performance Measures of the Tool. They are:

1. Accuracy in producing classification decisions (ratings and consumer advice) that are 'broadly consistent' with Australian community standards and classification decisions made by the Classification Board;
2. Ability to refuse classification to relevant material;
3. Ability to display classifications (ratings and consumer advice) produced by the Tool on the Australian Netflix web browser interface;
4. Ability to provide classifications (ratings and consumer advice) generated by the Tool in agreed format for display on the National Classification Database (NCD);
5. Ability to update the Australian Netflix Interface with new classification decisions as a result of revocations made by the Classification Board in a timely manner;
6. Satisfaction amongst Australian Netflix users (complaints management);
7. Ability to adapt to changes in the National Classification Scheme;

¹ Memorandum of Understanding between Netflix International B.V. and The Commonwealth of Australia

² Classification (Publications, Films and Computer Games) (Netflix Classification Tool) Approval 2016

8. Netflix's ability to make adjustments to the Tool to improve its performance;
9. The Tool must not produce a classification decision for films: already classified by the Classification Board / Classification Review Board / approved tool; public exhibition films; and/or upon advice from the Branch; and
10. Ability to advise the Director of the Board/Department of all Tool decisions.

2. Structure of the pilot

The pilot comprised the following broad components:

- The Tool was integrated with the classification database and the Netflix Australian user interface, to enable the display of Australian classification decisions generated by the Tool.
- To assess the performance and reliability of the Tool's classification decisions, in terms of being '*broadly consistent with Australian community standards and classification decisions made by the Classification Board*' and to identify areas for improvement, it was agreed that a sample of the Tool's classification decisions would be assessed. This activity was completed by trained Departmental assessors who viewed a film and made an independent recommendation to the Board about an appropriate classification rating and consumer advice³. The Board then made a determination on a classification rating and consumer advice for the film. This process is referred to in this report singularly as an *audit* and collectively as the *audit program*.

The audit program was conducted in two phases:

- In the first phase of the pilot, results of all audits were communicated to Netflix and used to improve the Tool. In addition, an agreed list of consumer advice was provided to Netflix to optimise consistency of consumer advice generated by the Tool to that applied by the Board.

s 22 - irrelevant to request

- In the second phase of the pilot, results from audits conducted on a random sample of film titles were analysed to assess whether the Tool produced classification ratings and consumer advice that were 'broadly consistent' with the classification decisions of the Board and Australian community standards.
- The pilot was evaluated by the Department against all of the agreed Key Performance Measures and the results presented in this report, s 22 - irrelevant to request

s 22 - irrelevant to request

³ Departmental assessors are trained in how to apply the Classification Act, the Code and the Film Guidelines when assessing content.

2.2. Audit program

To investigate and assess the performance of the Tool, an audit program consisting of two phases was established:

- Phase One – diagnostic phase (December 2016 – October 2017)
- Phase Two – evaluation phase (November 2017 to May 2018)

Trained Departmental assessors were responsible for assessing the films and providing a recommendation to the Board who then decided if the Tool's classification decision aligned with the classification decision of the Board for each title. The Board was provided with detailed reports and the opportunity to view the film which it used to inform its decision about whether classification decisions of the Tool aligned with classification decisions of the Board.

2.2.1. Phase One: Diagnostic phase – December 2016 to October 2017

In this phase, both targeted and random audits were conducted to identify issues affecting the alignment of the Tool's classification decisions with classification decisions of the Board. Feedback was given to Netflix to enable refinements to be made to the Tool. A small number of other titles were targeted for auditing owing to high media profile and/or complaints or enquiries. During this phase, 119 audits were conducted.

2.2.2. Phase two: Evaluation phase – November 2017 to May 2018

This phase focussed on assessing the Tool's *'ability to make decisions that are broadly consistent with Australian community standards and classification decisions made by the Classification Board'*.

Department completed 127 audits in the evaluation phase.

2.2.3. Definition of 'broadly consistent' decisions

The first of the Key Performance Measures, relating to the Tool's '*ability to make decisions that are broadly consistent with Australian community standards and classification decisions made by the Classification Board*' was given particular consideration because it establishes the standard for evaluating individual classification decisions of the Tool. It is crucial to the assessment of the Tool's reliability, that it produces consistent classification decisions across the range of classification categories⁶, which are consistent with decisions of the Board and Australian community standards.

The set of criteria for 'broadly consistent' classification decisions by the Tool was developed to ensure that consumers receive relevant information to inform their viewing choices and to effectively protect consumers from inappropriate, unsolicited or harmful content.

The formulation of consumer advice by the Board is determined by the most impactful of the classifiable elements⁷ and is governed by Board policy. An automated Tool will produce consumer advice that it has been programmed to generate. It is crucial that consumer advice provides useful information about the content of a film, so that it informs and assists consumers' choices.

As part of the evaluation criteria, it was decided that if the Tool generated a rating that was one rating *higher* than a rating decision of the Board, *and* that rating could be considered to be a 'borderline' decision, the Tool's rating would be evaluated as being a classification decision that is 'broadly consistent' with the Board and Australian community standards. However, a rating one or more categories *lower* than that which would have been given by the Board would *not* be deemed to be 'broadly consistent' because such a rating would not provide adequate protection to consumers.

⁶ The Australian classification categories are G, PG, M, MA 15+, R 18+, RC

⁷ The classifiable elements are themes, violence, sex, language, drug use and nudity.

Therefore, the definition of 'broadly consistent' agreed between the Branch and Netflix for the purposes of Phase Two of the auditing program (evaluation) was as follows:

- The rating is the *same* as the rating of the Classification Board; or
- The rating is one level *higher* than the rating of the Classification Board, where the rating could be considered 'borderline';

AND

- The classification decision contains an advisory from the agreed consumer advice list provided to Netflix and the content is present in the film.

Remainder of report (pp 10-18) s 22 - irrelevant to request