

## RESEARCH ARTICLE

# Elephant 'selfies': Evaluating the effectiveness of Instagram's warning of the potential negative impacts of photo opportunities with wild animals

Lauren A. Harrington<sup>1\*</sup>, Angie Elwin<sup>2</sup>, Neil D'Cruze<sup>1,2</sup>

**1** Wildlife Conservation Research Unit (WildCRU), Department of Biology, The Reccanati-Kaplan Centre, University of Oxford, Tubney, Oxfordshire, United Kingdom, **2** World Animal Protection UK, London, United Kingdom

\* [lauren.harrington@biology.ox.ac.uk](mailto:lauren.harrington@biology.ox.ac.uk)

## Abstract

Wildlife tourist attractions offering opportunities to observe, touch, and interact with wild animals, are visited by millions of people every year. Wildlife tourism has considerable economic value in many countries and can have positive impacts on wild animal populations (e.g. through habitat protection); it can also have negative impacts on population conservation and individual welfare (due to, e.g. habitat encroachment, disturbance, or disease). The recent phenomenon of 'wildlife selfies' shared on social media may seem harmless but can involve animals illegally or unsustainably captured from the wild, kept in poor conditions, or subject to cruel treatment. To address this issue, Instagram introduced a pop-up alert system that is triggered when users search for wild animal selfie hashtags (e.g. #elephantselfie), warning of the potential negative impacts of wildlife selfies on wild animals. Using elephant selfies as a case study, we found that Instagram's alert was triggered by only 2% of 244 elephant selfie-related hashtags tested. By comparing three pairs of similar hashtags (one of each pair that triggered the warning and one that did not), we were unable to detect a consistent difference in the type of post using each of the hashtags, the popularity of posts, or the sentiment of viewer comments. The warning is not shown when posting an image, or if a post is viewed directly by a follower, only if the post is encountered via a hashtag search. Currently, what is portrayed on social media appears to be inconsistent with apparent recent shifts in social acceptability regarding tourism, particularly as concerns direct contact between tourists and elephants. Instagram's wildlife selfie initiative was commendable but given its apparent lack of effect, we urge Instagram and other social platforms to do more to prevent harmful content from being posted on their platforms and to promote fair, ethical and sustainable interactions between wild animals and people.

## OPEN ACCESS

**Citation:** Harrington LA, Elwin A, D'Cruze N (2023) Elephant 'selfies': Evaluating the effectiveness of Instagram's warning of the potential negative impacts of photo opportunities with wild animals. *PLoS ONE* 18(4): e0283858. <https://doi.org/10.1371/journal.pone.0283858>

**Editor:** Bogdan Cristescu, Cheetah Conservation Fund, Namibia University of Science and Technology, NAMIBIA

**Received:** June 27, 2022

**Accepted:** March 17, 2023

**Published:** April 6, 2023

**Copyright:** © 2023 Harrington et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its [Supporting information](#) files.

**Funding:** This work was carried out in collaboration with, and funded by, World Animal Protection (WAP). Researchers from WAP listed as co-authors participated in study design, data collection and analysis, decision to publish, and preparation of the manuscript.

**Competing interests:** This study was funded by an animal welfare organisation; two of the authors are employed by the same organisation (NDC holds the position of Head of Research). Our intent in carrying out this study was to provide insight that would help inform recommendations and advice for Instagram (and other social media platforms) with respect to effectiveness of an initiative implemented by the platform to address potential harm associated with the phenomenon of wildlife selfies. Our results pertaining to observations on the appearance of the warning, and viewer response to posts, were in no way influenced by either the funding source, or our own personal views on animal welfare. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

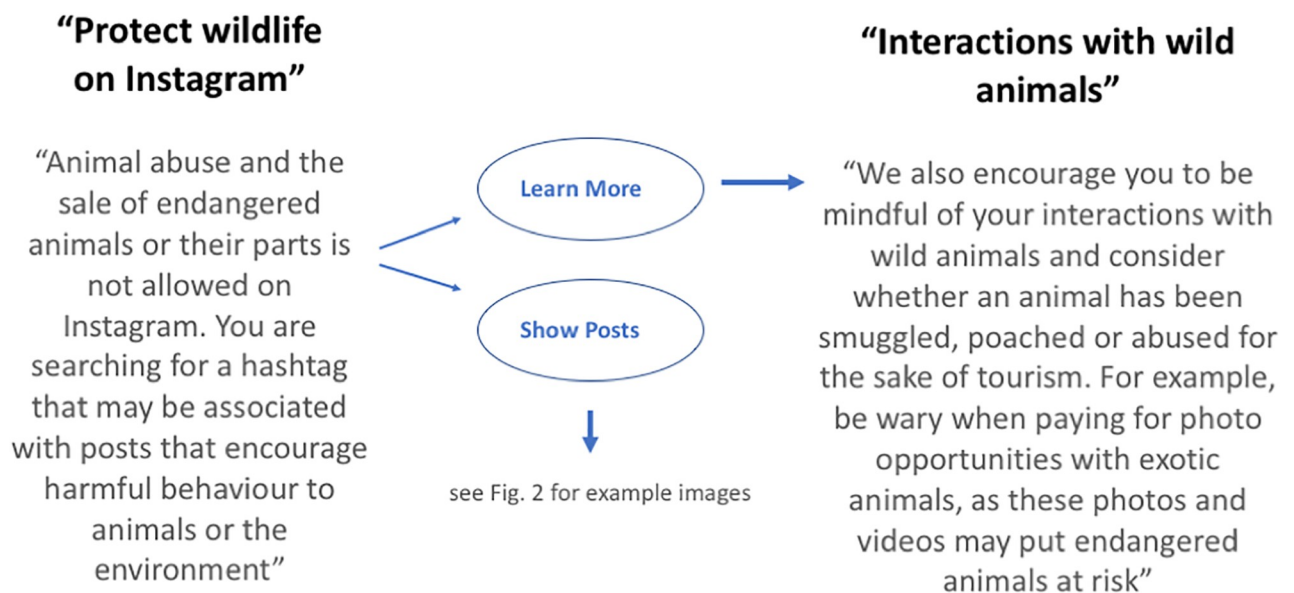
## Introduction

More than a billion tourists travel around the world every year (1.5 billion in 2019, [1]). Prior to global travel restrictions due to COVID-19 in 2020 and 2021, the number of international tourists was increasing at an annual rate of between three and seven percent [1]. Many people travel specifically to see wildlife, and for several countries (particularly developing countries) wildlife tourism is the leading foreign exchange earner [2–4]. Whilst some tourists and travellers desire what might be considered a more natural experience and are content to catch a fleeting glimpse of an elusive wild animal from a distance through binoculars, others increasingly demand guaranteed sightings (e.g. [5, 6]), want to be close to wild animals (e.g. [7]), and sometimes even touch them (e.g. [8]). Wildlife tourism can have significant benefits for wild animals (for example, through the support of protected areas and maintenance of important habitats), but there can also be negative impacts associated with human encroachment and disturbance (including, for example, disturbance of feeding and breeding behaviour, and disease transmission) [9–12]. Consequently, there may be a conflict between tourists and tourism operators and the wildlife that the tourists want to see [2].

Driven by peoples' (tourists) desire to be close to wild nonhuman animals (hereafter animals), wildlife tourist attractions (WTAs) around the world have become hugely popular [13, 14]. WTAs offer tourists and other visitors opportunities to interact with nondomestic animals, in captivity or in the wild. Globally, WTAs offer everything from walking with lions, swimming with dolphins, and feeding crocodiles to visiting sea turtle and civet coffee farms, viewing rescued orangutans, and watching dancing macaques. Like many other aspects of modern human life, these experiences are widely shared on social media (Instagram, Facebook, and others), increasingly as 'selfies' (where a person holds out a camera phone and photographs themselves [15, 16]). Here we are concerned with a particular type of selfie—'wildlife selfies'—in which photographers capture themselves with a wild animal in the same frame. The popularity of wildlife selfies is such that a number of species found in popular tourist destinations are used as part of a photo-prop industry where tourists pay to pose with them for photographic souvenirs. Examples include the common caiman, *Caiman crocodilus crocodilus*, and green anaconda, *Eunectes murinus*, in Brazil, [17], brown-throated three-toed sloth, *Bradypus variegatus*, in Brazil and Peru, [17, 18], slow loris, *Nycticebus* spp., in Thailand, [19], and barbary macaque, *Macaca sylvanus*, in Morocco, [20]. Wildlife selfies are presumably well intentioned by the person taking the photo, and may appear to be harmless, but tourist attractions offering wildlife selfie opportunities can have negative impacts on local populations of the species involved, and on the welfare of the individual animals (c.f. [13]), depending on where the animals are sourced, the conservation status of the species, and how they are treated. Animals used as photo props may be captured from the wild (e.g. [17, 21–23]), physically restrained when they are not on-show, kept in poor conditions (e.g. [17]) and, for the safety of the human handlers, subject to having their claws or teeth removed or clipped (practices that can result in death of the animal [19]). Visitors are often largely unaware of such negative impacts [13], and the taking and sharing of such images in a public domain (such as on social media) may promote desire among those that view the images to attend such attractions themselves [24]. Zoos and animal rescue centres also often offer opportunities for close encounters with captive-bred individuals of wild animals (as 'ambassadors'), presented by a trained handler in a setting that is usually intended to raise awareness and to build support for species protection. But without context these images cannot always be distinguished from those where animals are poorly treated and/or may have been taken illegally from the wild. There are concerns that these images may also propagate the photo-prop industry and encourage desire for inappropriate and/or illegal exotic companion animals [25–27].

Consequently, a number of organisations and groups advise against taking, and particularly, sharing such images on social media regardless of their provenance (e.g. [28, 29]).

In an attempt to protect captive wild animals from mistreatment, and to raise awareness of the potential conservation and welfare issues associated with tourist attractions offering photo opportunities with wild animals, Instagram introduced a pop-up alert system (Fig 1) that is triggered when users search for wild animal selfie hashtags using Instagram's search tool. Hashtags are words or phrases preceded by the symbol # that are used to tag content uploaded to social media. For example, '#slothselfie' might be used by someone posting an image of a selfie photograph with a sloth, which then enables any other user searching for selfies with sloths to find that particular post. The aim of the initiative was not to block access to the posts but to educate users on the welfare issues that may underlie a seemingly benign animal encounter [30]. The warning provided also includes a link to further information about wildlife exploitation (Fig 1) to help users understand more about the problem. Instagram's wildlife selfie alert initiative was implemented in December 2017 (more than 5 years ago to date). However, thus far (as far as we are aware) there has been no formal assessment of the policy's effectiveness. In the long-term, the intended effect of the warning might conceivably (and usefully) have been to discourage the posting of wildlife selfies and thus to prevent or reduce imitation behaviours and the taking of wildlife selfies in the first place. These type of long-term behaviour changes are hugely challenging to monitor and the relevant data are rarely available (c.f. [31]). However, a more immediate (and measurable) effect might be expected on the perceptions and attitudes of the viewers that are presented with the warning prior to viewing a wildlife selfie post. We hypothesized for example that an informed viewer having read the warning might respond more negatively to the post (i.e. be less likely to 'like' a post, or more likely to express negative sentiments in the comments). Although there is also a risk that such messaging has the unintended consequence of stimulating greater interest in wildlife selfie posts (the so-called Streisand effect [32]) we considered this unlikely for these type of posts (in contrast with, for example, a desire to watch violent content, c.f. [33]).



**Fig 1.** Instagram's wildlife selfie pop-up alert system showing the warning presented when users search for wild animal selfie hashtags (e.g. #elephantselfie) (left), the follow-on options given (middle), and an extract from the information provided (right) when users select 'Learn More'.

<https://doi.org/10.1371/journal.pone.0283858.g001>

We used Instagram posts depicting people in direct contact with elephants, a charismatic and popular species, and one of the most numerous captive-kept wild animals in WTAs [13], as a case study. The training required to ensure that tourists interacting with elephants are safe is described as being particularly cruel [34] and the conditions that elephants are kept in have been found, in many cases, to be inadequate in terms of animal welfare. Elephants used in tourist attractions are sometimes kept on chains, on concrete ground, and in the absence of social contact with other elephants [34, 35]. Riding elephants may cause elephants injury [36] but is still offered in some countries. Elsewhere, facilities offer tourists the opportunity to wash elephants in the river, to touch and to 'hug' their trunks, and to 'cuddle' juvenile elephants. These activities tend to be perceived as an ethical alternative to riding but can cause stress to elephants (e.g. [37, 38]) and are still dependent on the same type of training of young animals necessary to ensure (for the safety of visitors) that they can be controlled by their handler [35, 39].

A preliminary scoping study revealed that Instagram's wildlife selfie alert was triggered inconsistently among posts and so we were able to compare viewer response to posts using a hashtag that triggered the warning with those that used hashtags that did not trigger the warning. We posed the following research questions:

1. What proportion of hashtags related to elephant selfies trigger Instagram's warning?
2. Comparing elephant selfie posts that used hashtags that did trigger the warning with those that did not, is there any difference in:
  - a. post content
  - b. popularity of posts
  - c. perceptions or attitudes of viewers (as revealed by their comments)?

Overall, our aim was to assess how widely, and how effectively, the Instagram alert system is implemented and how viewers' response to posts might be influenced by it. Ultimately, we sought to inform refinements of this initiative as well as the design and implementation of other such social media initiatives intended to benefit the welfare of wild animals, and the status of their populations, through encouraging appropriate relationships between people and the wild animals that they interact with.

## Methods

### Definition of an elephant selfie

For the purpose of this study, we defined an 'elephant selfie' as an image of a person photographed in direct contact with an elephant, or standing in close proximity to an elephant (Fig 2), where 'direct contact' meant either touching, feeding, washing, or riding an elephant, and 'close proximity' was considered to be approximately within 'touching distance'. We included photographs taken as a 'selfie' by the person shown in the image, and photographs taken by someone else (the latter may also be referred to as a 'two shot image' [26]). This definition of wildlife selfies differs from the common meaning of the term 'selfie' (where the photo is taken by the person in the photo themselves) but follows standard usage and understanding of the term 'wildlife selfie' (as in, for example, [29]).

Our focus was specifically on visitors to wildlife tourist attractions, so we excluded images in circus settings, and those where the person in the image was clearly an elephant keeper or carer (indicated by the uniform worn, or the text accompanying the post). We also excluded images showing wild elephants in the background of the image (i.e. where the person was in





**Fig 2.** Example Instagram screenshots showing 'elephant selfie' images using the hashtags #elephantselfie (a, b), and #elephantride (c).

<https://doi.org/10.1371/journal.pone.0283858.g002>

an open safari vehicle), or images that had been digitally manipulated to make it appear that a person was close to an elephant, because these did not depict people in direct contact with elephants.

### Triggering of Instagram's wildlife selfie alert

To determine how often, and how effectively, Instagram's wildlife selfie alert is triggered, we initially searched Instagram, using the platform's search tool on the Explore page, for as many hashtags as we could find associated with elephant selfies using words related to our definition (e.g. #elephantselfie, #elephantselfies, #eleselfie, #elephantwashing, #elephantrides, and others) and recorded whether a subsequent search for posts under each identified hashtag triggered the warning. Additional relevant hashtags were obtained by snowball sampling [40], using all

hashtags suggested by Instagram under each term entered in the search bar. We included hashtags in the final count only if they led to one or more 'elephant selfie' posts (as defined above), and excluded those that were associated exclusively with posts by organisations (e.g. animal welfare groups) advocating against close interactions with elephants (e.g. #elephantridings-cruel), where images were included only for informative purposes. To account for frequency and recency of use, we recorded post volume (the number of posts that each hashtag had been used in) obtained from the search results provided by Instagram, and if the term had been used in posts posted since 2017 (i.e. since the introduction of Instagram's wildlife selfie alert). Post volumes were sometimes reported by Instagram only as approximate values so statistical comparison was not possible. All searches were completed between 28.09.21 and 20.10.21.

### Comparison of elephant selfie posts on Instagram using hashtags that do and do not trigger the warning

To explore in more detail the implementation, and effect of, Instagram's wildlife selfie alert we used a comparative study design in which we compared posts, and viewer response to posts, between pairs of similarly worded hashtags, one of which triggered the warning, and one of which did not. Our sample of hashtags was not random but was selected specifically to include terms that triggered the warning, that had a comparable (similarly worded) hashtag that did not trigger the warning, and that were used in a large number of Instagram posts. This design meant that inferences were limited to the specific hashtags used in the study but allowed any differences detected to be attributed to the (potential) triggering of the warning (rather than to differences in the terms used) and ensured a sufficient number of posts (and comments to posts) for estimation of post metrics and sentiment scores (below). Three pairs of similar hashtags were suitable for comparison: #elephantselfie and #elephantselfies, #elephantride and #elephantrides, #elephanthugs and #elephantcuddles; in each case, the first of each pair triggered the warning and the second did not (see Table 1).

**Post content.** To assess whether Instagram's wildlife selfie alert, on average, captures the content that it was intended to, we compared post content in relation to the presence/absence of hashtags that triggered the warning. For each hashtag, two independent observers viewed and categorised posts shown on a mobile device (in the order presented on the device, under 'top' where the option was available). Each post shown was counted as being within one of six pre-defined categories, as defined in Table 1. These categories were intended to broadly describe the situation of the elephant in the image and the presence/role of the people in the image, and to be representative of the type of animal-tourist images that might be seen on social media generally. Viewing was terminated when most posts were dated 2017 or earlier,

**Table 1. Categories used to describe posts.** 'Situation' refers to whether the elephant is captive (C) or wild (W); the presence of people in the image is denoted by Y (yes) or N (no), and the role of the people in the image as either 'in close proximity to, or touching the elephant' (close/touching) or 'observing from a distance' (observing).

Category	Description	Situation	Presence/role of people
Elephant selfie	an image of a person photographed in direct contact with an elephant, or standing in close proximity to an elephant (see text)	C	Y (close, touching)
Captive elephant	captive-held elephants photographed alone, i.e. painted elephants, or elephants clearly in a zoo or with a keeper or carer	C	Y or N
Elephant safaris	wild elephants shown with people in the foreground of the image, i.e. in an open safari vehicle	W	Y (observing)
Wild elephants	wild elephants pictured in natural habitats, with no people in the image	W	N
Advocacy	elephant selfie images included in a post advocating against direct interactions with elephants	C	Y (close, touching)
Irrelevant	all other images, e.g. toy elephants, drawings of elephants, or any other image not showing a live elephant	NA	NA

<https://doi.org/10.1371/journal.pone.0283858.t001>

were categorised as 'irrelevant', or (in the case of #elephantride/s) when a minimum of 400 posts had been viewed.

Because posts are not necessarily shown in the same order on different devices or under different user accounts, we used chi-squared tests to test for broad agreement between observers in the relative distribution of content categories under each hashtag rather than tests of inter-observer agreement based on paired observations. This approach means that any difference detected may be due *either* to differences between observers (lack of observer reliability or agreement) *or* inconsistency between devices or Instagram accounts. In all cases, we detected no difference between observers (chi-squared test,  $p > 0.08$  in all cases; six tests performed separately, Bonferroni-corrected critical  $p$  value = 0.008).

**Viewer response.** To explore the effect of the use of hashtags that triggered the warning on viewer response to elephant selfie posts specifically, we, first, quantified the popularity of posts, and, second, analysed viewer comments on the posts using sentiment analysis. For this analysis, we used only those posts that conformed to our definition of an elephant selfie (above) and that had been posted in 2018 or later. For comparative purposes, posts that used more than one of the six hashtags were excluded. For each of the six hashtags, we manually collated data for each post including the date that it had been posted, the country where it was posted (obtained from the geo-location provided, or inferred from the text contained within the post), and the number of likes that the post received. Still image posts and video posts were treated in the same way and we did not distinguish between them. We used the number of likes as an indicator of post popularity to reflect a combination of the number of times a post was viewed *and* liked; this is not the same as engagement rate (calculated as the number of likes as a percentage of the number of followers) which measures popularity *within* the audience that ordinarily follow the posters account. Both metrics have weaknesses. The disadvantage of the number of likes as a measure of popularity is that it provides no information on the number of viewers that did not like the post. Whereas engagement rate does not account for the number of views received via searches for hashtags and so may significantly overestimate engagement for posts using popular hashtags. On balance, and mindful of data accessibility and privacy issues (and the fact that the number of followers can only be obtained by accessing personal accounts), we considered number of likes most suitable for the purposes of this study. Posts were retrieved in the order in which they were presented on a desktop computer, and the search for relevant posts continued until either there were no more posts available, or most were irrelevant or dated 2017 or earlier (up to a maximum of approximately 100 posts). For the hashtags that triggered a warning, only a limited number of posts were shown on a desktop computer, in which case data for the remaining posts were manually copied from a mobile device. Our sample of posts, thus represents a subset of all possible posts but includes the 'top' (the most highly viewed) and the most recent posts shown.

Finally, viewer comments were manually copied from each post, and comments on all posts combined in a single document for each of the six hashtags. Documents were saved as plain text files for analysis, where lines of text corresponded to individual comments. For this analysis we were only able to retrieve comments from posts shown on a desktop computer (see above). We quantified the sentiment of comments to posts using each of the six hashtags, using a sentence-level lexicon-based approach. A sentence-level approach assigns a sentiment score to each sentence (comment) based on the sum of the scores of all meaningful (scorable) words within the comment (i.e. all words that exist in the lexicon and have an associated score, taking account of the word count used). We used the "sentimentr" package [41] in R to assign sentiment scores, which is able to account for valence shifters and modifiers in the text and allows the incorporation of emojis. Emojis enhance, and modify the meaning of, the text [42] and so their inclusion not only adds information from the comments based only on emojis,

but also improves the accuracy of sentiment scoring compared to using the linguistic text alone [43, 44]. Prior to analysis names, dates, profile pictures, identities, personal comments irrelevant to the context of the image (e.g. comments referring to personal circumstances or appearance), advertising (e.g. of other local sanctuaries), and the word 'instagram', were manually identified and removed. Foreign text was replaced with English language using google translate where possible. Emojis were separated into individual symbols so that each counted as a single 'word', and replaced with identifiers that can be recognised by sentimentr, using the "textclean" [45] package in R. The sentimentr package provides a sentiment score between -1 and +1 for each scorable word, and includes an emoji lexicon (with comparable scores for each listed emoji, based on [42])—we combined these two into a new bespoke lexicon and obtained sentiment scores for each comment (based on the text and emojis in each comment) using the 'sentiment\_by' function in sentimentr. Mean sentiment scores were calculated for the complete comment text for each of the six hashtags by averaging the individual sentence scores, and standard errors of the means estimated using the package "plotrix" [46]; in addition, the number of negative, neutral, and positive sentence sentiment scores were counted for each hashtag. In addition, to provide context for sentiment scores, we identified the most frequently occurring words in the comment text (including emojis as words) for each of the six hashtags, using the text mining package "tm" [47, 48]. Stopwords were excluded, except for the word 'not' due to its potential importance as a valence shifter; where not was identified as a frequently occurring word we examined the original comment text to determine the context that it was used in.

### Ethical considerations

Our study involved covert observation [49]; however, we used only data (images shown, post metrics, and comments) that were publicly available on the Instagram platform, and posted on public accounts. No automated data collection methods were used, and data were collected and analysed within the terms and conditions of Instagram ([www.instagram.com](http://www.instagram.com)). We did not engage in deceptive practice, and did not engage directly or otherwise with Instagram users (either those posting images or commenting on images). This was an independent study that was not formally linked with an academic institution; therefore, none of the authors had access to an institutional ethics review body. In the absence of formal external ethical review we relied on published guidelines and principles for the use of social media data in research. It is noteworthy that this is a rapidly evolving field, there is currently a lack of *standardised* guidance, review boards often do not have appropriate expertise, and there is disagreement amongst authors publishing on this topic (see e.g. [50], and references therein). Bearing this in mind, we took two standpoints. Firstly, to protect the identity of individual Instagram users (and thus to ensure anonymity), user names were not recorded, and user names, profile pictures, and any names and/or personal details were removed from text files created and saved for analysis (this is in accordance with ethical research practices as outlined in [51] and [52]). Secondly, because commenters were not identified individually, and data deriving directly from comments (with the exception of four anonymised quotes used for illustrative purposes, section 3.4) are presented here only in aggregate form (single words and sentiment scores), we considered that informed consent was not necessary (this is in accordance with [52]).

### Statistical analysis

All statistical analyses were carried out in R (version 4.1.2; [53]). For left-skewed data, we report medians, and use a robust coefficient of variation ( $RCV_Q$ , based on the median and the interquartile range  $[IQR/median * 0.75]$  [54]) as a measure of relative dispersion within



samples. We used Pearson's chi-square test to test for a difference in post content between posts that used hashtags that triggered Instagram's alert system and those that used hashtags that did not trigger the alert, for an association between the origin (country) of a post and the hashtag used, and for differences in the proportion of positive, negative, neutral comments between pairs of hashtags. We compared popularity (quantified as the number of likes) between each pair of hashtags using non-parametric two sample Mann-Whitney U Tests, and used Welch's two sample t tests to test for a difference in mean sentiment scores between pairs of terms. In all cases, pair-wise comparisons were carried out separately and Bonferroni corrections used to account for multiple tests. Effect sizes were quantified, for non-parametric tests, using Hodges-Lehmann (HL) estimator (which is the sample median of all cross-sample pairwise differences) and, for parametric tests, using the mean difference; we used non-standardised effect sizes because these are straightforward to interpret for the parameters (number of likes and sentiment scores) used. Violin plots, used to depict the distribution and range of the data, were drawn using the "vioplot" package in R [55].

## Results



### How many hashtags related to elephant selfies trigger Instagram's warning?

We identified a total of 244 hashtags associated with Instagram posts that included various types of elephant selfie (Table 2). Three-quarters of the hashtags identified ( $n = 183$ ) had been used since 2017. When searched for, 2% of all hashtags identified ( $n = 5$ ; 2.7% of those that had been used since 2017) triggered Instagram's warning. Post volumes for hashtags that triggered the warning were  $> 1,000$  in all cases (maximum = 156,169, Table 2). Post volumes for hashtags that did not trigger the warning were hugely variable (range 1 –  $> 10,000$ ,  $n = 239$ , Table 2, S1 Appendix) such that whilst 159 (66.5%) hashtags that did not trigger the warning had post volumes  $< 10$ , 18 (7.5%) had post volumes  $> 1,000$ , and two (0.8%) had post volumes  $> 10,000$ .

### Comparison of elephant selfie posts on Instagram using hashtags that do and do not trigger the warning

We initially viewed a total of 2,512 Instagram posts in order to describe post content in relation to the presence/absence of hashtags that trigger the warning ( $n = 468$  using the term #elephantselfie<sup>1</sup>, 225 #elephantselfies, 871 #elephantride<sup>1</sup>, 487 #elephantrides, 193 #elephanthugs<sup>1</sup>, and 268 #elephantcuddles; where the terms marked with an exclamation mark triggered Instagram's warning and those without did not). Analyses of post popularity were carried out on a subset of viewed posts that conformed to our definition of elephant selfies (above). Thus our main dataset comprised a sample of 440 elephant selfie posts, posted between 2018 and 2021, that used at least one of the six hashtags ( $n = 57$  used the term #elephantselfie, 48 #elephantselfies, 122 #elephantride, 16 #elephantrides, 97 #elephanthugs, and 96 #elephantcuddles, excluding four posts that used more than one of the hashtags). Location could be inferred for 403 (91.6%) of elephant selfie posts (S2 Appendix). Elephant selfie posts originated from 15 different Asian and African countries, and the USA, but were predominantly from Thailand (50% posts,  $n = 218$ ), India (14.2%,  $n = 62$ ), and Indonesia (11.9%,  $n = 52$ ). Posts from India were most likely to be associated with #elephantride or #elephantrides (82.3% of India origin posts,  $n = 51$ ), whereas those from Thailand (51.8% of Thai origin posts,  $n = 113$ ) and Indonesia (59.6% Indonesian origin posts,  $n = 31$ ) were most likely to be associated with #elephanthugs or #elephantcuddles (association between country and hashtag:  $\chi^2 = 93.70$ ,  $df = 6$ ,  $p < 0.001$ ,

**Table 2. Hashtags associated with elephant selfie posts on Instagram.** Data shown are the post volume (the number of posts in which the hashtag was used), whether Instagram's pop-up alert was triggered (Y = yes, N = no), and whether any of the posts had been posted (i.e. the hashtag had been used) since the warning was introduced in 2017. Not all posts conformed to our definition of an elephant selfie; hashtags were included if at least one post using it was relevant (see footnote). Only hashtags with a post volume > 500 are shown here, the full set of identified relevant hashtags is provided in [S1 Appendix](#). Neither the list of hashtags here, or the full list of hashtags given in [S1 Appendix](#), is intended to be exhaustive. Bold text indicates hashtags used for further analysis.

Search term	Post volume	Warning triggered	Used in posts since 2017
#elephantselfie	22269	Y	Y
#elephantselfies	705	N	Y
#elephantride	82601	Y	Y
#elephantrides	16922	N	Y
#elephantriding	29952	Y	Y
#elephantsanctuary	156169	Y	Y
#selfiewithelephant	654	N	Y
#elephantsanctuaryphuket	5884	N	Y
#elephantbath	12752	N	Y
#elephantsanctuarychiangmai	9436	N	Y
#elephantsanctuarythailand	1538	N	Y
#elephantride 	1000+	N	Y
#elephantriders	1000+	N	Y
#elephantrider	1000+	N	Y
#elephantriding 	1000+	N	Y
#elephantsanctuarybles	1000+	N	Y
#elephantbathing	5000+	N	Y
#elephantbathtime	500+	N	Y
#elephantwashing	1000+	N	Y
#bathingwithelephants	1000+	N	Y
#eleselfie	500+	N	Y
<b>#elephanthugs</b>	<b>1000+</b>	<b>Y</b>	<b>Y</b>
<b>#elephantcuddles</b>	<b>1000+</b>	<b>N</b>	<b>Y</b>
#feedingelephants	5000+	N	Y
#elephantfeeding	5000+	N	Y
#bathingelephants	1000+	N	Y
#washingelephants	1000+	N	Y

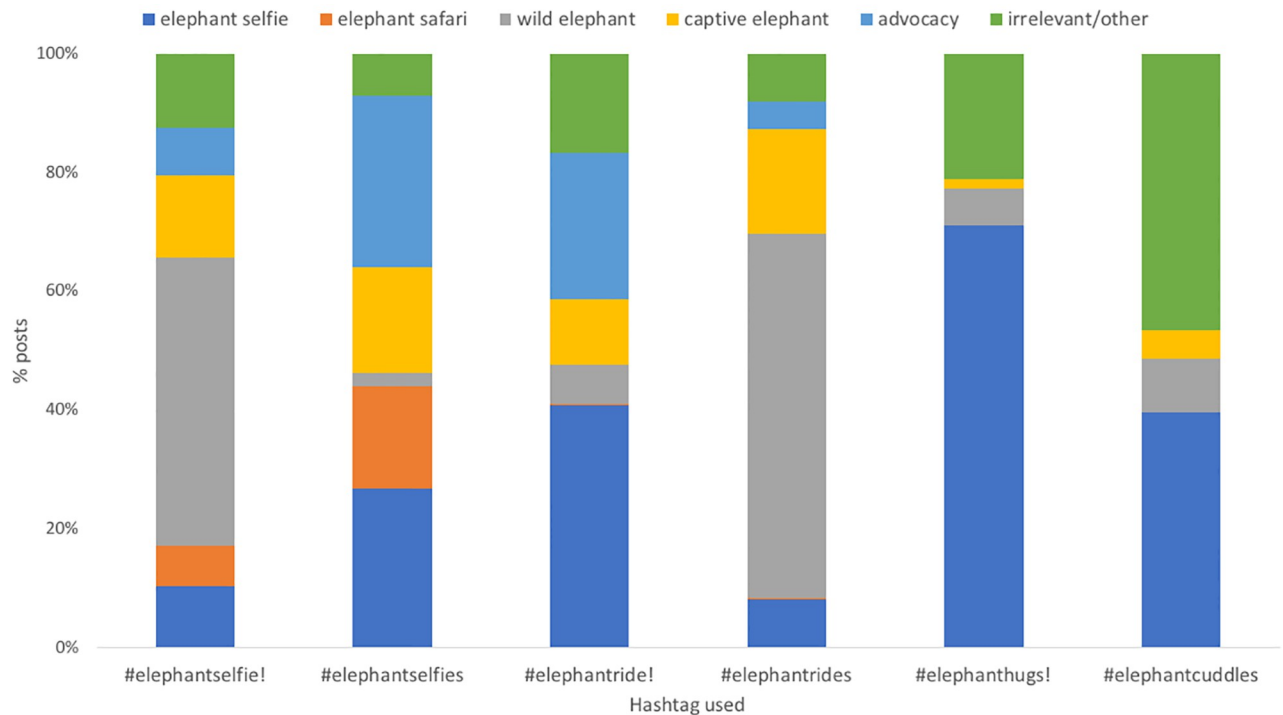
The following hashtags were used exclusively in posts advocating against elephant rides or direct contact with elephants: #elephantridesarecruel, #elephantridingistorture, #elephantridingsucks, #elephantridingisnotcool, #elephantridingisabuse, #elephantridingiswrong, #elephantridingiscruel, #elephantrides✘.

<https://doi.org/10.1371/journal.pone.0283858.t002>

countries combined as Thailand, India, Indonesia, 'Other', and hashtags combined as #elephantride/s, #elephantselfie/s, #elephanthugs/cuddles). Because viewer comments were only retrieved from posts shown on a desktop computer (see [Methods](#)), sentiment scores were calculated for a reduced dataset comprising a total of 221 posts (n = 12 using the term #elephantselfie, 48 #elephantselfies, 27 #elephantride, 16 #elephantrides, 21 #elephanthugs, and 96 #elephantcuddles).

### How does post content differ among posts using hashtags that do and do not trigger the warning?

For the pair of terms #elephantride and #elephantrides, posts using #elephantride (which did trigger the warning) comprised a larger proportion of 'elephant selfie' posts (40.6%, n = 354 of 871 posts) than did posts using #elephantrides (8.0%, n = 39 of 487) (which did not trigger the

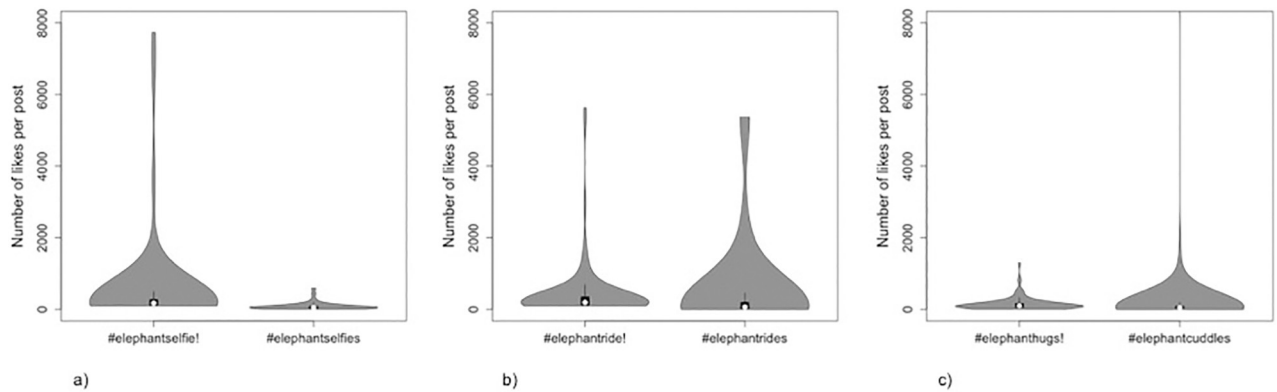


**Fig 3. Types of post using each of six hashtags, where the first of each 'pair' (#elephantselfie-#elephantselfies, #elephantride-#elephantrides, #elephanthugs-#elephantcuddles) triggers Instagram's warning (marked by an!) and the second does not.** Relative proportions of post types were statistically significantly different between pairs of terms in all cases;  $n = 468, 225, 871, 487, 193, 268$ , respectively. Content categories defined in Methods and Table 1.

<https://doi.org/10.1371/journal.pone.0283858.g003>

warning); the latter comprised predominantly (61.4%,  $n = 299$  of 487) 'wild elephant' posts (Fig 3). A similar trend in elephant selfie posts was apparent between the terms #elephanthugs (which did trigger the warning) and #elephantcuddles (which did not): posts using #elephanthugs comprised 71.0% 'elephant selfie' posts ( $n = 137$  of 193 posts), whilst posts using #elephantcuddles comprised 39.6% 'elephant selfie' posts ( $n = 106$  of 268). For #elephantcuddles, the majority of posts using the term were categorised as 'irrelevant' (46.6%,  $n = 125$  of 268 posts; we did not attempt to describe this category in any further detail but noted that several posts showed images of soft toy elephants, some photographed alongside human babies). The pair of terms #elephantselfie and #elephantselfies, however, showed the opposite trend with posts using #elephantselfie (which did trigger the warning) comprising a smaller proportion of 'elephant selfie' posts (10.3%,  $n = 48$  of 468 posts) than did posts using #elephantselfies (26.7%,  $n = 60$  of 225) (which did not trigger the warning); in this case, the posts using the term that triggered the warning (#elephantselfie) comprised predominantly (48.5%) 'wild elephant' posts ( $n = 227$  of 468). Advocacy posts comprised a relatively large proportion (28.9 and 24.8%, respectively) of posts using the hashtags #elephantselfies and #elephantride (terms that did not, and did, trigger the warning, respectively).

Overall, all six hashtags were used in all types of post, with the exception of #elephanthugs and #elephantcuddles, which were not used in either 'elephant safari', or 'advocacy' posts. However, although there were statistically significant differences in the relative proportions of types of post between pairs of hashtags (chi-squared tests,  $p < 0.001$  in all cases, Bonferroni-corrected critical  $p$  value for 3 tests = 0.017), there was no consistent pattern between those that did trigger Instagram's warning and those that did not (Fig 3).



**Fig 4.** Violin plots showing popularity (as measured by number of likes) of elephant selfie posts using each of six hashtags, where the first of each pair of terms—a). #elephantselfie and #elephantselfies, b). #elephantride and #elephantrides, c). #elephanthugs and #elephantcuddles—triggers Instagram's warning (marked by an!) and the second does not. Circles depict the median ( $RCV_Q = 0.64, 0.88, 0.86, 1.94, 0.91, 0.97$ , respectively); maximum value for #elephantselfie not shown to improve visual comparison.

<https://doi.org/10.1371/journal.pone.0283858.g004>

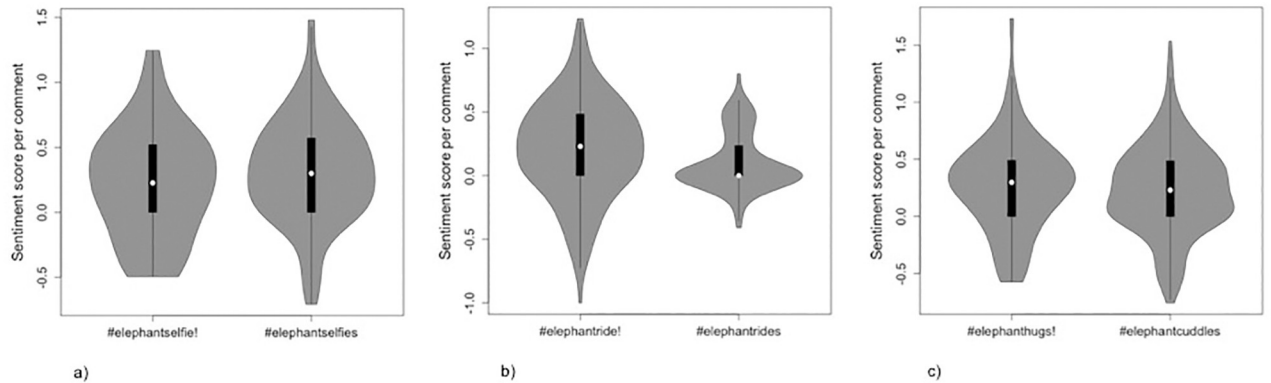
### Is there any difference in the popularity of 'elephant selfie' posts using hashtags that do and do not trigger the warning?

Popularity of elephant selfie posts (measured as the number of likes) was highly variable within and between hashtags, and heavily-skewed towards lower values (i.e. most posts received relatively few likes, Fig 4). Median number of likes across hashtags ranged between 51.5 and 194.5, whilst maximum values for five of the six hashtags exceeded 1,000 ( $> 24,000$  for posts using #elephantselfie, Fig 4); for all hashtags, within-term robust coefficients of variation were relatively high ( $RCV_Q = 0.64\text{--}1.94$ , summary data in S3 Appendix). The number of likes for posts using hashtags that triggered Instagram's warning was statistically significantly different from the number of likes for those that did not, for all pairs of terms ( $p < 0.01$ , in all cases; Bonferroni-corrected critical p value for 3 tests = 0.017), but, contrary to expectation, median likes for posts using hashtags that triggered the warning exceeded median likes for those that did not (i.e. were more popular), in all cases. Median cross-sample pairwise differences (HL estimators) were 112, 105, and 38 for #elephantselfie/s, #elephantride/s, and #elephanthugs/cuddles, respectively (i.e., on average, a post using the hashtag 'elephantselfie' received 112 more likes than did a post using the hashtag 'elephantselfies').

### Is there any difference in the perceptions or attitudes of viewers (as revealed by their comments) to posts using hashtags that do and do not trigger the warning?

For posts using all hashtags, sentiment scores of comments were approximately normally distributed and ranged between a minimum of  $-1.0$  to  $-0.41$  and a maximum of  $0.8$  to  $1.73$  (i.e. all contained both negative and positive comments). Mean sentiment scores were positive in all cases, with values between 0.11 and 0.32. Mean sentiment scores did not differ significantly between comments to posts using #elephantselfie (that triggered the warning) and those using #elephantselfies (that did not trigger the warning;  $p = 0.165$ ), or between those using #elephanthugs and #elephantcuddles (that did and did not trigger the warning, respectively;  $p = 0.397$ , Fig 5, Table 3); however, mean sentiment for comments to posts using #elephantride was statistically significantly higher (more positive) than for comments to posts using





**Fig 5.** Violin plots showing the distribution and range of sentiment scores for comments on elephant selfie posts using each of six hashtags, where the first of each 'pair' of terms—a). #elephantselfie and #elephantselfies, b). #elephantride and #elephantrides, c). #elephanthugs and #elephantcuddles—triggers Instagram's warning (marked by an!) and the second does not. Circles depict the median, the bars depict the inter-quartile range.

<https://doi.org/10.1371/journal.pone.0283858.g005>

#elephantrides (i.e. comments to posts that used a hashtag that triggered the warning were more positive than those that did not;  $t = 2.83$ ,  $df = 237$ ,  $p = 0.005$ , Bonferroni-corrected critical  $p$  value for 3 tests = 0.017, mean difference = 0.10, Fig 5, Table 3). There was also a statistically significant association between the polarity of the sentiment score and whether or not the hashtag used triggered the warning for comments to posts using #elephantride/s ( $p < 0.001$ ; but not for those using #elephantselfie/s or #elephanthugs/cuddles,  $p = 0.119$ , 0.093, respectively, Bonferroni-corrected critical  $p$  value for 3 tests = 0.017)—although in this case, comments to posts using the hashtags that did trigger the warning (#elephantride) appeared to have more negative *and* positive scores (i.e. more polarised scores) rather than more positive scores specifically (there was no association between the presence of the warning and the proportion of negative and positive comments [ $p = 1.0$ ], but there was a statistically significant association between the presence of a warning and the proportion of polarised and neutral comments [ $p < 0.001$ ]).

The most frequently occurring word (or symbol) in the comments to posts for all hashtags was a 'smiley face with heart eyes', and the second or third most frequent a heart symbol (Table 4). Overall, for all hashtags, words and symbols tended to be positive in nature (including, for example, the words 'love', 'amazing', 'beautiful'), with some exceptions for comments to posts using the two hashtags associated with riding elephants (#elephantride and #elephantrides). However, symbols depicting more negative emotions (such as anger ['angry red face'] and sadness ['crying face']) occurred in comments to both elephant riding hashtags regardless of whether or not they triggered Instagram's warning (Table 4). The same was true of the word 'not' which also occurred in comments to posts using both elephant riding hashtags, and was used, in more than half of all occurrences ( $n = 11$  of 20 occurrences, and 7 of 10 occurrences in comments to #elephantride and #elephantrides, respectively) in phrases like "...tourists benefit, *not* the elephants", "should *not* be encouraged", "*not* your entertainers", "*not* a pet", "*not* here to serve us".

## Discussion

Instagram and other social media platforms (such as YouTube) are currently entirely self-regulated [56]. As such, user-generated content posted on most social media platforms is usually subject to some form of acceptable standards or community guidelines derived by each

**Table 3. Sentiment scores for comments to posts using each of six hashtags, where the first of each ‘pair’ of terms triggers Instagram’s warning and the second does not.** Sentiment scores (assigned, based on text and emojis combined, using the “sentimentr” package (Rinker 2021) in R), reported as mean, SE, and median, and percentage negative, neutral, and positive; sample size (n) is given as the number of posts, the total number of comments for all posts using each of the hashtags combined, and the number of words used in the analysis.

#	Warning	n posts	n	Sentiment score	Sentiment score
			comments, words	mean (SE), median	(%) neg, neut, pos
#elephantselfie	Y	12	49, 217	0.220 (0.060), 0.226	24.5, 6.1, 69.4
#elephantselfies	N	48	108, 718	0.319 (0.038), 0.299	13.0, 13.0, 74.0
#elephantride	Y	27	248, 2487	<b>0.206 (0.024)*, 0.229</b>	<b>23.4, 6.9, 69.7**</b>
#elephantrides	N	16	81, 660	<b>0.107 (0.025)*, 0.000</b>	<b>11.1, 56.8, 32.1**</b>
#elephanthugs	Y	21	83, 483	0.271 (0.044), 0.299	21.7, 4.8, 73.5
#elephancuddles	N	96	384, 2700	0.230 (0.020), 0.231	18.7, 13.3, 68.0

\* statistically significantly different (p < 0.01);  
 \*\* statistically significantly different (p < 0.001).

<https://doi.org/10.1371/journal.pone.0283858.t003>

individual platform (for Instagram, see <https://help.instagram.com>). Compliance, or indeed detection of non-compliance, is, in most cases, at least partly dependent on viewers identifying and flagging inappropriate content (e.g. [57]). In the case of wildlife selfies this is difficult because there is nothing in the image itself to suggest cruelty to nonhuman animals or negative conservation impacts on threatened species; neither is it the case that every image of a smiling human depicted alongside a wild animal is associated with detrimental impacts on animals at either an individual, population, or species level. To understand these issues requires much greater detail on where individual wild animals depicted are sourced from, the conditions that

**Table 4. Most frequently occurring words in comments on elephant selfie posts using each of six hashtags, where the first of each ‘pair’ of terms—a). #elephantselfie and #elephantselfies, b). #elephantride and #elephantrides, c). #elephanthugs and #elephancuddles—triggers Instagram’s warning (marked by an!) and the second does not.** Data are the frequency of occurrence of words and/or emojis appearing the comments, including only those with a frequency of occurrence of 10 or more. Stopwords were not included, except for the word ‘not’ which may be important as a valence shifter. Words and emojis with negative sentiment scores are shaded in the table.

#elephantselfie <sup>1</sup>	#elephantselfies	#elephantride <sup>1</sup>	#elephantrides	#elephanthugs <sup>1</sup>	#elephancuddles
29	39	97	60	45	114
	16	49	28	17	56
	16	beautiful 26	21		love 29
		not 20	13		omg 17
		15	11		amazing 17
		14	11		16
		12	10		cute 15
		12	not 10		12
		12	10		beautiful 11
		12			10
		wow 11			
		amazing 11			
		love 10			
		10			

<https://doi.org/10.1371/journal.pone.0283858.t004>

they are kept in (e.g. [17]), and the training that they are subject to, than is possible from a single image, as well as a much deeper understanding of their welfare requirements in captivity (e.g. [58]), the welfare impacts associated with handling [18], and the conservation status of the species in the wild (e.g. [21]). In this respect, Instagram should be commended for launching an initiative that attempts to educate users on some of the more subtle (or otherwise hidden) issues associated with the use of photo props in tourist destinations [30, 59] (see also [13, 60]) and of wild animal exploitation more broadly. However, to be genuinely informative, educational messages need to be visible to the correct target audience, easy to understand, sufficiently repetitive that they are noticed and remembered, consistent (to avoid ambiguity or confusion), relevant, and, in addition, users need to be receptive to the message [61–64].

Three key findings of this study suggest that Instagram's wildlife selfie alert system as currently implemented is limited, inconsistent, and, in many cases, appears to be mismatched with the type of content that it was intended to highlight. Amongst 244 hashtags related to elephant selfies that we were able to identify, 98% failed to trigger Instagram's alert system. Differences in hashtag terms that did and did not trigger the warning were extremely minor; for example, #elephantselfie did trigger the warning, #elephantselfies did not, yet both were associated with posts showing the types of wildlife selfie identified as being of potential concern (i.e. those showing close or direct contact with wild animals, cf. [65]). Conversely, #elephantselfie (which did trigger the warning) was associated, in almost half of all posts examined, with images showing only apparently wild elephants where the warning was irrelevant. The other two pairs of hashtags examined (#elephantride and #elephantrides, and #elephanthugs and #elephantcuddles) similarly failed to show any apparent pattern in associated content that might explain why the warning was triggered for one and not the other. Whilst some of the hashtags identified that did not trigger the warning were infrequently used, and our study was limited to elephant selfies, a brief search for other wildlife species featured in wildlife selfies on Instagram reveals similar inconsistencies: #koalaselfie (with 5,000+ posts), #slothselfie (22,600 posts), and #monkeyselfie (14,100 posts) all trigger the warning but the plural versions of these terms—#koalaselfies (100+ posts), #slothselfies (500+ posts), #monkeyselfies (1,000+ posts)—and a variety of other related terms, do not. Further, whilst not part of our initial study, *post hoc* observations revealed that the warning is triggered only when hashtags are used to search for images and is not triggered when posting images or when following a user account, meaning that users can view and interact with ('like' or comment on) posts without ever encountering the warning. Ultimately the warning does not prevent content from being posted.

The words and emojis used most frequently in post comments (Table 3), together with mean comment sentiment scores (which were  $> 0$  in all cases, Table 2), suggest that elephant selfie posts included in the study were perceived as loveable, exciting, visually appealing, overall positive, and enviable experiences regardless of the hashtag used by the poster. Although, for each of the hashtags, there was some level of negativity in post comments (Table 2, Fig 5), there was no evidence that this was associated with the presence of Instagram's warning. The effect sizes observed for popularity of posts (where posts using hashtags that triggered the warning appeared to be more popular than those using hashtags that did not trigger the warning) were small relative to the range of values recorded and likely not meaningful. That a statistically significant difference between mean comment sentiment scores (albeit also with a small mean difference) was observed only for posts using hashtags associated with elephant riding (where posts with the hashtag that triggered the warning received more positive comments on average than those with the hashtag that did not trigger the warning) suggests that this was an artefact of sampling rather than an effect of the warning per se. It is noteworthy, for example, that we were only able to include 16 posts that used the term #elephantrides. Further, the occurrence of some clearly negative words and emojis amongst those most frequently used in

post comments, appeared to be associated with the act of riding elephants rather than presence of the warning. Overall, post comments suggested an apparent lack of broader awareness of the potential negative welfare impacts on elephants involved in alternative activities ('hugging', 'cuddling', and washing young and/or adult elephants), despite travel guidelines advising against such activities [66].

There are complexities associated with the long-tradition of elephant keeping in South East Asia (historically, for logging) in terms of the fate of those elephants already in captivity, and the livelihoods of their keepers (e.g. [67]) that are beyond the scope of this study. Nevertheless, our observations of social media posts, and viewer response to them, highlight inconsistencies between what is portrayed on social media and current trends in tourism reflecting apparent shifts in social acceptability. Over 200 travel companies and tour operators, for example, have committed to move away from selling or promoting venues offering elephant shows, rides, or other forms of direct contact between tourists and elephants [68]. Further research on other types of WTA in terms of how they are portrayed, and responded to, on social media in relation to current animal welfare guidelines (and, for threatened species, conservation assessments) is required to determine whether this is the case more broadly.

This is the first study of which we are aware that has attempted to assess the effect of Instagram's wildlife selfie alert. There are few, if any, directly comparable initiatives on which to draw lessons. Although there is a body of research on 'trigger warnings' or 'content warnings' (e.g. [69–71]), these types of warnings are designed to warn viewers of potentially distressing content (i.e. to protect the viewer themselves), a fundamentally different objective to the selfie alert that aims to protect the animal in the image, other wild animals in a similar situation, and populations of the species shown. The use of counter-statements or warnings to prevent the spread of misinformation on social media is comparable to the wildlife selfie alert system in some respects (although the nature, and object, of harm associated with misinformation differs). One recent study in this area [72] was able to show that an unverified health-related tweet was less likely to be retweeted when it was accompanied by a counter-statement or warning. We were not able to monitor sharing of Instagram posts but we found no evidence that posts using a hashtag that triggered the warning were either less popular, or were perceived by commenters as more negative, than those that used hashtags that did not trigger the warning.

Our study had some limitations. Most relevant is that we were unable to assess how many people actually saw Instagram's warning or, crucially, how many saw the warning and subsequently did not go on to view the associated posts. This has implications for the interpretation of post comments because our sample of comments will have been made either by users who were not shown the warning or by users who chose to ignore the warning. However, whilst this type of potential sampling bias might have influenced user response for those posts that used hashtags that triggered the warning (and could perhaps explain the apparent, albeit relatively minor, popularity effects observed), it does not negate our findings of rather limited, and inconsistent, triggering of the warning. Further, whilst the fact that the warning is only presented to users searching for hashtag terms has implications for our study design, it also represents a weakness in the implementation of Instagram's warning. Indeed, posts that do not use any hashtags would neither trigger the warning or be detected in our initial search, highlighting further weaknesses in the system and also suggesting that our sample of posts might underestimate the prevalence of wildlife selfies on Instagram. In terms of study limitations, it is also possible that our failure to detect a statistically significant difference in sentiment scores between posts using hashtags that did and did not trigger the warning (for two of the pairs of hashtags) was due, in part, to limited sample size (number of comments and thus words available for analysis) and insufficient precision in estimated scores. There is little guidance on recommended sample size in sentiment analysis, but [73]'s analysis demonstrates increasing



precision with increasing number of words up to 300 to 400 words, with little gain in precision for sample sizes over one to two thousand words. In our study, the number of words used in the estimation of sentiment scores varied between 217 and > 2,500 but was < 300 for only one of the six hashtags, so we tentatively conclude that sample size was sufficient. There are also interpretation issues in assigning sentiment scores. One example in this case is the word 'jealous', which is scored as negative, but in this context presumably means that the viewer likes what they see and is jealous of the person in the photo. Similarly, the phrase "no riding elephants" or "you should not ride elephants" is given a sentiment score of 0, but is not, in this context, a neutral comment, rather it is a negative response to a photo of a person riding an elephant. The use, and intended meaning, of specific emojis may also vary with demographic, psychological, and/or cultural differences among users [74, 75], and may be used ironically [43]. These issues are rarely given explicit attention in sentiment analysis studies using social media data, but mean that average sentiment scores, and even overall polarity, may be inaccurate (e.g. [44, 76]). In this case, however, whilst the precise numerical sentiment scores reported should be interpreted with caution, it is the comparison between pairs of hashtags, rather than the scores themselves, that is relevant to the questions posed in this study. Finally, whilst our sample of hashtags was not randomly sampled, and we cannot strictly extrapolate our findings beyond the particular hashtags used in this study, consistency in average post popularity and comment sentiment scores observed *between* pairs of terms (i.e. between #elephantselfie/s and #elephantride/s, Figs 4 and 5) together with the range of terms used (selfie, ride, hugs/cuddles) suggest that our sample of hashtags is probably representative of other non-similar hashtags used.

## Conclusion

In conclusion, whilst Instagram's wildlife selfie alert system is well-intentioned and has some value, it needs to be implemented more widely, consistently, and triggered at different points in the user process, to achieve its intended purpose. For elephants specifically, there is growing emphasis among operators on responsible tourism [66, 68, 77]; currently, content posted on Instagram does not reflect this. Inadequacies associated with Instagram's current focus on specific hashtags suggest that alternative approaches, such as image recognition might be more effective. Image recognition is already used in detecting serious crime online [78, 79], and the technology is well developed for identifying species in camera trap images (e.g. [80]). The International Fund for Animal Welfare in collaboration with Baidu (a Chinese technology company specialising in internet services) recently launched an artificial intelligence (AI)-powered tool to identify images of endangered wildlife products traded online [81], and World Animal Protection used image recognition alongside key word searches and machine learning to train social listening algorithms to identify inappropriate wildlife selfies online [82]. AI technology is used by a number of social media platforms, including Instagram and Facebook (both owned by Meta), to detect and remove content that violates platform standards or guidelines (e.g. [83]); currently, however, although Instagram prohibits coordination of poaching of endangered species and the sale of any live animals, Instagram's Community Guidelines make no mention of inappropriate wild animal content, or of wildlife selfies specifically [84]. Sponsored advertisements posted on social media might also be more effective in informing people of the potential hidden harm to wild animals involved in animal selfie opportunities for tourists and the photo prop industry. In terms of assessing the success of this or any other related initiative (c.f. [85]), Instagram might also consider a mechanism whereby users can provide feedback to the warning itself to provide a measure of exposure (how many people actually see and read it), interest, and perceived relevance, which would in turn inform

refinements and/or further educational initiatives as part of a co-design approach [86]. Further detailed analysis of wildlife selfie posts on social media, and viewer response to them, is needed, but preliminary observations suggest that our findings, and suggestions, are relevant not only to elephants used in tourism but to many other species commonly involved in wildlife selfies. Overall, more needs to be done by the major social media platforms to ensure that content permitted on social media (across platforms) does not involve, or is not associated with, threat to wild populations, animal cruelty, or other types of interactions with wildlife that are deemed detrimental to their welfare or otherwise breach community guidelines. Instagram and other large social media platforms undoubtedly have considerable resources and technological expertise at their disposal. We urge them to use these wisely in order to play their part in promoting healthy, fair, ethical and sustainable interactions between wild animals and growing human populations.

## Supporting information

**S1 Appendix. Expanded Table 1 showing all hashtags that we were able to identify associated with elephant selfie posts on Instagram.** Data shown are the post volume (the number of posts in which the hashtag was used), whether Instagram's pop-up warning was triggered (Y = yes, N = no), and whether any of the posts had been posted (i.e. the hashtag had been used) since 2017 (since the warning had been introduced). Note that not all posts using each hashtag conformed to the definition of elephant selfie used in the study (or were relevant to the aims of the study) but hashtags were included if at least one post was relevant (see footnote).

(DOCX)

**S2 Appendix. Country locations for 440 elephant selfie posts on Instagram using the hashtags #elephantselfie, #elephantselfies, #elephantride, #elephantrides, #elephanthugs, #elephantcuddles.**

(DOCX)

**S3 Appendix. Popularity (number of likes) of Instagram posts using each of six hashtags, where the first of each pair of terms triggers Instagram's warning and the second does not.** IQR = inter-quartile range,  $RCV_Q = (\text{quartile-based}) \text{ robust coefficient of variation} = [IQR / \text{median}] * 0.75$ ; HL-estimator is Hodges-Lehmann estimator (or the sample median of all cross-sample pairwise differences).

(DOCX)

## Acknowledgments

The authors would like to thank Nick Stewart, Stephanie Chen, and Morvah George at World Animal Protection for initial discussions that assisted the development of this research. We also thank Andrew Harrington for help with categorising posts, and Alix Harrington for preparation of Fig 2.

## Author Contributions

**Conceptualization:** Lauren A. Harrington, Angie Elwin, Neil D'Cruze.

**Data curation:** Lauren A. Harrington, Angie Elwin.

**Formal analysis:** Lauren A. Harrington.

**Funding acquisition:** Neil D'Cruze.

**Investigation:** Lauren A. Harrington, Angie Elwin, Neil D’Cruze.

**Methodology:** Lauren A. Harrington.

**Project administration:** Neil D’Cruze.

**Writing – original draft:** Lauren A. Harrington.

**Writing – review & editing:** Lauren A. Harrington, Angie Elwin, Neil D’Cruze.

## References

1. United Nations World Tourism Organization. Global and regional tourism performance. 2022 [accessed 24th January 2022] <https://www.unwto.org/global-and-regional-tourism-performance>
2. Reynolds PC, Braithwaite D. Towards a conceptual framework for wildlife tourism. *Tour Manag.* 2001; 22: 31–42.
3. Twining-Ward L, Li W, Bhammar H, Wright E. Supporting Sustainable Livelihoods through Wildlife Tourism. Washington, DC: World Bank Group; 2018.
4. Travel World & Council Tourism. The Economic Impact of Global Wildlife Tourism—August 2019. London, UK: [www.wtfc.org](http://www.wtfc.org); 2019.
5. Kubo T, Mieno T, Kuriyama K. Wildlife viewing: The impact of money-back guarantees. *Tour Manag.* 2019; 70: 49–55. <https://doi.org/10.1016/j.tourman.2018.06.010>
6. Knight J. The ready-to-view wild monkey: The Convenience Principle in Japanese Wildlife Tourism. *Ann Tour Res.* 2010; 37: 744–762. <https://doi.org/10.1016/j.annals.2010.01.003>
7. Van Hamme G, Svensson MS, Morcatty TQ, Nekaris KAI, Nijman V. Keep your distance: Using Instagram posts to evaluate the risk of anthroponotic disease transmission in gorilla ecotourism. *People and Nature.* 2021; 3: 325–334. <https://doi.org/10.1002/pan3.10187>
8. Dwyer J. Do Not Feed the Animals: Do Not Touch: Desire for Wild Animal Companionship in the Twenty-first Century. *Inter Studies Lit Environ.* 2012; 19: 623–644.
9. Roe D, Leader-Williams N, Dalal-Clayton B. Take Only Photographs, Leave Only Footprints: The Environmental Impacts of Wildlife Tourism. IIED Wildlife and Development Series, No.10; 1997.
10. Orams MB. Feeding wildlife as a tourism attraction: a review of issues and impacts. *Tour Manag.* 2002; 23: 281–293. [https://doi.org/10.1016/S0261-5177\(01\)00080-2](https://doi.org/10.1016/S0261-5177(01)00080-2)
11. Tapper R. United Nations Environment Programme (UNEP) & the Convention on Migratory Species (CMS) Secretariat. *Wildlife Watching and Tourism: A Study on the Benefits and Risks of a Fast Growing Tourism Activity and its Impacts on Species.* Bonn, Germany: UNEP/CMS Secretariat; 2006.
12. Green RJ. Disturbing Skippy on Tour: Does It Really Matter? Ecological and Ethical Implications of Disturbing Wildlife. In: *Wildlife Tourism, Environmental Learning and Ethical Encounter.* Springer; 2017: pp. 221–233.
13. Moorhouse TP, Dahlsjö CAL, Baker SE, D’Cruze NC, Macdonald DW. The Customer Isn’t Always Right—Conservation and Animal Welfare Implications of the Increasing Demand for Wildlife Tourism. *PLoS One.* 2015; 10: e0138939. <https://doi.org/10.1371/journal.pone.0138939> PMID: 26489092
14. Moorhouse T, D’Cruze NC, Macdonald DW. Unethical use of wildlife in tourism: what’s the problem, who is responsible, and what can be done? *J Sustainable Tourism.* 2017; 25: 505–516.
15. Miltner K, Baym N. The Selfie of the Year of the Selfie: Reflections on a Media Scandal. *Int J Comm.* 2015; 9: 1701–1715.
16. Weilenmann A, Hillman T. Selfies in the wild: Studying selfie photography as a local practice. *Mobile Media & Communication.* 2020; 8: 42–61. <https://doi.org/10.1177/2050157918822131>
17. D’Cruze N, Machado F, Matthews N, Balaskas M, Carder G, Richardson V, et al. A review of wildlife ecotourism in Manaus, Brazil. *Nat Conserv.* 2017; 22: 1–16. <https://doi.org/10.3897/natureconservation.22.17369>
18. Carder G, Plese T, Machado FC, Paterson S, Matthews N, McAnea L, et al. The Impact of ‘Selfie’ Tourism on the Behaviour and Welfare of Brown-Throated Three-Toed Sloths. *Animals.* 2018; 8: 216. <https://doi.org/10.3390/ani8110216> PMID: 30463203
19. Osterberg P, Nekaris KAI. The use of animals as photo props to attract tourists in Thailand: A case study of the slow loris *Nycticebus* spp. *TRAFFIC Bulletin.* 2015; 27: 13–18.
20. Stazaker K, Mackinnon J. Visitor Perceptions of Captive, Endangered Barbary Macaques (*Macaca sylvanus*) Used as Photo Props in Jemaa El Fna Square, Marrakech, Morocco. *Anthrozoös.* 2018; 31: 761–776.

21. D'Cruze N, Niehaus C, Balaskas M, Vieto R, Carder G, Richardson VA, et al. Wildlife tourism in Latin America: Taxonomy and conservation status. *J Sustainable Tourism*. 2018; 26: 1562–1576.
22. Nijman V. An Assessment of the live elephant trade in Thailand. Cambridge, UK: TRAFFIC International; 2014. <https://www.traffic.org/site/assets/files/8423/elephant-trade-thailand.pdf>
23. Prakash TGSL, Indrajith WAADU, Aththanayaka AMCP, Karunarathna S, Botejue M, Nijman, et al. Illegal capture and internal trade of wild Asian elephants (*Elephas maximus*) in Sri Lanka. *Nat Conserv*. 2020; 42: 51–69. <https://doi.org/10.3897/natureconservation.42.57283>
24. van der Meer E, Botman S, Eckhardt S. I thought I saw a pussy cat: Portrayal of wild cats in friendly interactions with humans distorts perceptions and encourages interactions with wild cat species. *PLoS One*. 2019; 14: e0215211. <https://doi.org/10.1371/journal.pone.0215211> PMID: 31042719
25. Learmonth MJ. Human-Animal Interactions in Zoos: What Can Compassionate Conservation, Conservation Welfare and Duty of Care Tell Us about the Ethics of Interacting, and Avoiding Unintended Consequences?. *Animals*. 2020; 10: 2037. <https://doi.org/10.3390/ani10112037> PMID: 33158270
26. Spooner SL, Stride JR. Animal-human two-shot images: Their out-of-context interpretation and the implications for zoo and conservation settings. *Zoo Biology*. 2021; 40: 563–574. <https://doi.org/10.1002/zoo.21636> PMID: 34223658
27. Shaw MN, McLeod EM, Borrie WT, Miller KK. (2021). Human Positioning in Close-Encounter Photographs and the Effect on Public Perceptions of Zoo Animals. *Animals*. 2021; 12: 11. <https://doi.org/10.3390/ani12010011> PMID: 35011117
28. Waters S, Setchell JM, Maréchal L, Oram F, Wallis J, Cheyne SM. Best Practice Guidelines for Responsible Images of Non-Human Primates. A Publication of The IUCN Primate Specialist Group, Section for Human-Primate Interactions; 2021.
29. Ward-Paige CA. Even scientists take selfies with wild animals. Here's why they shouldn't. *The Conversation*. 2016; June 28.
30. Daly N. Exclusive: Instagram Fights Animal Abuse With New Alert System. *National Geographic*, 2017; December 4. <https://www.nationalgeographic.com/animals/article/wildlife-watch-instagram-selfie-tourism-animal-welfare-crime>
31. Veríssimo D, Wan AKY. Characterizing efforts to reduce consumer demand for wildlife products. *Conserv Biol*. 2019; 33: 623–633. <https://doi.org/10.1111/cobi.13227> PMID: 30259569
32. Bergman JN, Buxton RT, Lin H-Y, Lenda M, Attinello K, Hajdasz AC, et al. Evaluating the benefits and risks of social media for wildlife conservation. *FACETS*. 2021; 7: 360–397. <https://doi.org/10.1139/facets-2021-0112>
33. McCarthy-Jones S. From Tarantino to Squid Game: why do so many people enjoy violence? *The Conversation*. 2021; October 27.
34. Schmidt-Burbach J. Taken For a Ride. *World Animal Protection*; 2017.
35. Schmidt-Burbach J, Ronfot D, Srisangiam R. Asian Elephant (*Elephas maximus*), Pig-Tailed Macaque (*Macaca nemestrina*) and Tiger (*Panthera tigris*) Populations at Tourism Venues in Thailand and Aspects of Their Welfare. *PLoS One*. 2015; 10: e0139092. <https://doi.org/10.1371/journal.pone.0139092> PMID: 26407173
36. Kontogeorgopoulos N. The Role of Tourism in Elephant Welfare in Northern Thailand. *J Tour*. 2009; 10: 1–19.
37. Millsaugh JJ, Burke T, Van Dyk G, Slotow R, Washburn BE, Woods RJ. Stress Response of Working African Elephants to Transportation and Safari Adventures. *J Wildl Manag*. 2007; 71: 1257–1260. <https://doi.org/10.2193/2006-015>
38. Bansiddhi P, Brown JL, Khonmee J, Norkaew T, Nganvongpanit K, et al. Management factors affecting adrenal glucocorticoid activity of tourist camp elephants in Thailand and implications for elephant welfare. *PLoS One*. 2019; 14: e0221537. <https://doi.org/10.1371/journal.pone.0221537> PMID: 31574099
39. World Animal Protection. Cruel elephant training process 'the crush' exposed. 2020a [accessed 03.05.2022] [www.worldanimalprotection.org.uk](http://www.worldanimalprotection.org.uk).
40. Parker C, Scott S, Geddes A. Snowball Sampling. In: Atkinson P, Delamont S, Cernat A, Sakshaug JW, Williams RA, editors. *SAGE Research Methods Foundations*; 2019.
41. Rinker TW. sentimentr: Calculate Text Polarity Sentiment. R package version 2.9.0; 2021 <https://github.com/trinker/sentimentr>
42. Novak PK, Smailović J, Sluban B, Mozetič I. Sentiment of Emojis. *PLoS One*. 2015; 10: e0144296. <https://doi.org/10.1371/journal.pone.0144296> PMID: 26641093
43. Tian Y, Galery T, Dulcinati G, Molimpakis E, Sun C. Facebook sentiment: Reactions and Emojis. *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media, W17-1102*, Association for Computational Linguistics; 2017. <https://aclanthology.org/W17-1102/>



44. Gupta S, Singh A, Ranjan J. Sentiment Analysis: Usage of Text and Emoji for Expressing Sentiments. *Adv Data Informat Sci*. 2020; 94: 477–486. [https://doi.org/10.1007/978-981-15-0694-9\\_45](https://doi.org/10.1007/978-981-15-0694-9_45)
45. Rinker TW. textclean: Text Cleaning Tools. R package version 0.9.3; 2018. <https://github.com/trinker/textclean>
46. Lemon J. Plotrix: a package in the red light district of R. *R-News*. 2006; 6: 8–12.
47. Feinerer I, Hornik K, Meyer D. Text Mining Infrastructure in R. *J Stat Softw*. 2008; 25: 1–54. <https://doi.org/10.18637/jss.v025.i05>
48. Feinerer I, Hornik K. tm: Text Mining Package. R package version 0.7–6; 2018. <https://CRAN.R-project.org/package=tm>
49. Thompson RM, Hall J, Morrison C, Palmer NR, Roberts DL. Ethics and governance for internet-based conservation science research. *Conserv Biol*. 2021; 35: 1747–1754. <https://doi.org/10.1111/cobi.13778> PMID: 34057267
50. Samuel G, Buchanan E. Guest editorial: Ethical issues in social media research. *J Emp Res on Human Res Ethics*. 2020; 15: 3–11. <https://doi.org/10.1177/1556264619901215> PMID: 31959061
51. Zook M, Barocas S, Boyd D, Crawford K, Keller E, Gangadharan SP, et al. Ten simple rules for responsible big data research. *PLoS Comput Biol*. 2017; 13: e1005399. <https://doi.org/10.1371/journal.pcbi.1005399> PMID: 28358831
52. Townsend L, Wallace C. Social media research: a guide to ethics. University of Aberdeen: Workshop proceedings, Social Media, Privacy and Risk: Towards More Ethical Research Methodologies; 2016. <https://ahrecs.com/resources/>
53. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021. <https://www.R-project.org/>
54. Arachchige CN, Prendergast LA, Staudte RG. Robust analogs to the coefficient of variation. *J Appl Stat*. 2020; 49: 268–290. <https://doi.org/10.1080/02664763.2020.1808599> PMID: 35707217
55. Adler D, Kelly ST. vioplot: violin plot. R package version 0.3.7; 2020. <https://github.com/TomKellyGenetics/vioplot>
56. Esmail N, Wintle B, 't Sas-Rolfes M, Athanas A, Beale C, Bending Z, et al. Emerging illegal wildlife trade issues: a horizon scan. *Conserv Lett*. 2020; 13: e12715.
57. Harrington LA, Elwin A, Paterson S, D'Cruze N. The viewer isn't always right—response to fake animal rescues on YouTube and implications for social media self-policing policies. *People and Nature*. 2023; 5: 103–118. <https://doi.org/10.1002/pan3.10416>
58. Kitson H, Nekaris K. Instagram-fuelled illegal slow loris trade uncovered in Marmaris, Turkey. *Oryx*. 2017; 51: 394–394.
59. Eltagouri M. Taking a #koalaselfie or #dolphins photo? Instagram has something to say about it. *The Washington Post*, 2017; December 5. <https://www.washingtonpost.com/news/animalia/wp/2017/12/05/taking-a-koalaselfie-instagram-has-something-to-say-about-it/>
60. Nekaris KAI, Musing L, Vazquez AG, Donati G. Is Tickling Torture? Assessing Welfare towards Slow Lorises (*Nycticebus* spp.) within Web 2.0 Videos. *Folia Primatologica*. 2015; 86: 534–551. <https://doi.org/10.1159/000444231> PMID: 26848764
61. Suggs LS, McIntyre C, Warburton W, Henderson S, Howitt P. Communicating health messages: A framework to increase the effectiveness of health communication globally. *World Innovative Summit for Health*; 2015.
62. Maibach E. Increasing public awareness and facilitating behavior change: Two guiding heuristics. In: *Climate Change and Biodiversity*. New Haven, Connecticut: Yale University Press; 2017.
63. World Health Organisation. WHO Strategic Communications Framework for effective communications. Geneva, Switzerland: World Health Organisation; 2017.
64. Greszczuk C. Making messages work. The Health Foundation, [www.health.org.uk](http://www.health.org.uk); 2020. <https://www.health.org.uk/what-we-do/a-healthier-uk-population/thinking-differently-about-health/making-messages-work>
65. World Animal Protection. A close up on cruelty: The harmful impact of wildlife selfies in the Amazon. London: World Animal Protection; 2017.
66. ABTA. How the travel industry is working to improve the welfare of elephants in tourism. 2022 [accessed 08.04.22]. <https://www.abta.com/sustainability/animal-welfare/how-travel-industry-working-improve-welfare-elephants-tourism>
67. Duffy R, Moore L. Global regulations and local practices: the politics and governance of animal welfare in elephant tourism. *J Sustainable Tourism*. 2011; 19: 589–604.
68. World Animal Protection. Elephant-friendly travel companies. 2019 [accessed 08.04.2022]. [www.worldanimalprotection.org/wildlife-not-entertainers](http://www.worldanimalprotection.org/wildlife-not-entertainers)

69. Jones PJ, Bellet BW, McNally RJ. Helping or Harming? The Effect of Trigger Warnings on Individuals With Trauma Histories. *Clin Psychol Sci*. 2020; 8: 905–917. <https://doi.org/10.1177/2167702620921341>
70. Boysen GA, Isaacs RA, Tretter L, Markowski S. (2021). Trigger warning efficacy: The impact of warnings on affect, attitudes, and learning. *Scholarsh Teach Learn Psychol*. 2021; 7: 39–52. <https://doi.org/10.1037/stl0000150>
71. Charles A, Hare-Duke L, Nudds H, Franklin D, Llewellyn-Beardsley J, Rennick-Egglestone S, et al. Typology of content warnings and trigger warnings: Systematic review. *PLoS One*. 2022; 17: e0266722. <https://doi.org/10.1371/journal.pone.0266722> PMID: 35507551
72. Öztürk P, Li H, Sakamoto Y. Combating Rumor Spread on Social Media: The Effectiveness of Refutation and Warning. 2015; 48th Hawaii International Conference on System Sciences: 2406–2414.
73. Yazman J. How Much Text Do We Really Need for Sentiment Analysis? *RPubs*; 2017: 10/07/2017.
74. Bai Q, Dan Q, Mu Z, Yang M. A Systematic Review of Emoji: Current Research and Future Perspectives. *Front Psychol*. 2019; 10. <https://doi.org/10.3389/fpsyg.2019.02221> PMID: 31681068
75. Kutsuzawa G, Umemura H, Eto K, et al. Classification of 74 facial emoji's emotional states on the valence-arousal axes. *Sci Rep*. 2022; 12: 398. <https://doi.org/10.1038/s41598-021-04357-7> PMID: 35087076
76. Lennox RJ, Verissimo D, Twardek WM, Davis CR, Jarić I. Sentiment analysis as a measure of conservation culture in scientific literature. *Conserv Biol*. 2020; 34: 462–471. <https://doi.org/10.1111/cobi.13404> PMID: 31379018
77. World Animal Protection. Elephants. Not commodities. Taken for a Ride 2. 2020b [accessed 03.05.2022] [https://www.worldanimalprotection.org/sites/default/files/2021-06/Elephants\\_Not\\_commodities\\_Full\\_Report.pdf](https://www.worldanimalprotection.org/sites/default/files/2021-06/Elephants_Not_commodities_Full_Report.pdf)
78. Bursztein E, Clarke E, DeLaune M, Eliff DM, Hsu N, Olson L, et al. Rethinking the Detection of Child Sexual Abuse Imagery on the Internet. In: *The World Wide Web Conference*. New York: Association for Computing Machinery; 2019: pp. 2601–2607.
79. Pereira M, Dodhia R, Anderson H, Brown R. Metadata-Based Detection of Child Sexual Abuse Material. *arXiv*, 2021; 2010.02387v2.
80. Norouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer MS, Packer C, et al. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *PNAS*. 2018; 155 (25). <https://doi.org/10.1073/pnas.1719367115> PMID: 29871948
81. International Fund for Animal Welfare. AI Guardian of Endangered Species recognizes images of illegal wildlife products with 75% accuracy rate. Press Release, 22<sup>nd</sup> April 2020. <https://www.ifaw.org/press-releases/ai-endangered-species-recognize-images-illegal-wildlife>
82. World Animal Protection. Image recognition software reveals impact of cruel wildlife selfies on social media. Press Release, 5<sup>th</sup> October 2017. <https://www.worldanimalprotection.org/news/image-recognition-software-reveals-impact-cruel-wildlife-selfies-social-media>
83. Meta. How technology detects violations. Transparency Center, 19th January 2022. <https://transparency.fb.com/en-gb/enforcement/detecting-violations/technology-detects-violations/> [accessed 3<sup>rd</sup> November 2022]
84. Meta. Community Guidelines. Instagram Help Centre, 2022. <https://www.facebook.com/help/instagram/477434105621119>
85. Doughty H, Wright J, Verissimo D, Lee JSH, Oliver K, Milner-Gulland EJ. Strategic advertising of online news articles as an intervention to influence wildlife product consumers. *Conserv Sci Pract*. 2020; 2: e272. <https://doi.org/10.1111/csp2.272>
86. Bowie MJ, Dietrich T, Cassey P, Verissimo D. Co-designing behavior change interventions to conserve biodiversity. *Conserv Sci Pract*. 2020; 2:e278. <https://doi.org/10.1111/csp2.278>