

뉴스 기사 신뢰도



측정 방안:



뉴스 트러스트

사례를 중심으로

오세욱 송해엽

연구서 2020 - 03

뉴스 기사 신뢰도 측정 방안: 뉴스 트러스트 사례를 중심으로

책임연구 오세욱 한국언론진흥재단 선임연구위원
공동연구 송해엽 군산대학교 미디어문화학과 교수
보조연구 전창영 한국외국어대학교 미디어커뮤니케이션학과 박사

발행인 표완수

편집인 김철훈

발행일 2020년 11월 30일 초판 제1쇄 발행

한국언론진흥재단

주소 04520 서울특별시 중구 세종대로 124 프레스센터빌딩

전화 (02) 2001-7759

팩스 (02) 2001-7740

www.kpf.or.kr

편집 (주)케이에스센세이션

주소 07237 서울특별시 영등포구 국회대로66길 11, 102호(여의도동, 퍼스텍빌딩)

전화 (02) 761-0031

팩스 (02) 761-2727

인쇄 (사)한국장애인e-Work협회

주소 07333 서울특별시 영등포구 국제금융로 8길 19 중앙빌딩 411호

전화 (02) 783-0800

팩스 (02) 783-0750

이 책에 실린 내용은 본 재단의 공식 견해가 아닌 필자의 연구 결과임을 밝힙니다.

© 한국언론진흥재단, 2020

ISBN 978-89-5711-593-0 93070

이 보고서는 정부광고 수수료로 조성된 언론진흥기금의 지원을 받아 제작되었습니다.

자세한 내용은 QR코드를 통해 확인해 주세요.



CONTENTS

| | |
|-------------------------------|-----|
| 01. 문제의 제기 | 1 |
| 02. 계량적 기준에 따른 뉴스 기사 배열 사례 검토 | 7 |
| 1. 이론적 논의 | 7 |
| 2. 계량적 기준에 따른 뉴스 기사 배열 | 12 |
| 3. 자동 배열 사례들의 한계 | 25 |
| 03. 뉴스 트러스트 위원회 알고리즘 사례 분석 | 27 |
| 1. 뉴스 트러스트 위원회의 출범과 경과 | 27 |
| 2. 뉴스 트러스트 위원회 논의 및 개발 과정 | 35 |
| 3. 뉴스 트러스트 알고리즘 개발 결과 | 64 |
| 4. 뉴스 트러스트 알고리즘의 한계와 의미 | 92 |
| 04. 뉴스 기사 신뢰 요인에 관한 이용자 인식조사 | 94 |
| 1. 조사 설계 | 94 |
| 2. 조사 결과 | 98 |
| 3. 소결 | 144 |
| 05. 결론 및 제언 | 150 |

LIST OF TABLES

| | | |
|----------|--------------------------------|-----|
| [표 2-1] | 국내 상황에서 저널리즘이 수행하는 기능 | 8 |
| [표 2-2] | 세부차원에 따른 저널리즘 품질요소 | 9 |
| [표 2-3] | 뉴스 기사의 품질 측정에 사용된 품질 요소와 평가 방식 | 10 |
| [표 2-4] | 크레더(credder)의 기사 평가 기준 | 15 |
| [표 3-1] | 뉴스 트러스트 위원회 1기 주요 경과 | 32 |
| [표 3-2] | 뉴스 트러스트 위원회 2기 주요 경과 | 33 |
| [표 3-3] | 뉴스 트러스트 위원회 3기 주요 경과 | 34 |
| [표 3-4] | 측정 가능한 요인 초안 | 45 |
| [표 3-5] | 위원회에서 논의한 측정 요인 안 | 48 |
| [표 3-6] | 분류별 기사 길이에 관한 기초통계 | 53 |
| [표 3-7] | 평가에 관한 합의사항 | 59 |
| [표 3-8] | 저널리즘 가치와 계량 요인 관계 기반 가중치 안 | 62 |
| [표 3-9] | 저널리즘 가치와 계량 요인 관계 기반 가중치 결과 | 63 |
| [표 3-10] | 기사의 길이 분포 통계 | 77 |
| [표 3-11] | 11개 계량 요인별 기본 점수 부여 방식 | 85 |
| [표 3-12] | 공식 1에 따른 예시 기사의 기본 점수 | 88 |
| [표 3-13] | 공식 2에 따른 예시 기사의 저널리즘 가치 점수 | 89 |
| [표 3-14] | 공식 3에 따른 예시 기사의 저널리즘 가중치 점수 | 90 |
| [표 4-1] | 설문조사 참여자 | 97 |
| [표 4-2] | 신뢰하는 기자와 언론사 명단 | 115 |

LIST OF FIGURES

| | | |
|-----------|-----------------------------|-----|
| 〈그림 2-1〉 | 딥뉴스 다이제스트 | 18 |
| 〈그림 2-2〉 | 다음카카오 루빅스의 기사 배열 과정 | 21 |
| 〈그림 2-3〉 | 조선닷컴 기사 하단 데이블의 자동 추천 콘텐츠 | 23 |
| 〈그림 3-1〉 | 위원회에서 논의한 측정 요인 안 | 51 |
| 〈그림 3-2〉 | 대학원생이 평가한 기사 품질과 기사 길이의 관계 | 54 |
| 〈그림 3-3〉 | 기사 품질과 심층성에 대한 개별 위원 평가결과 | 61 |
| 〈그림 3-4〉 | 뉴스 트러스트 전체 개발 흐름도 구상 | 65 |
| 〈그림 3-5〉 | 뉴스 트러스트 전체 개발 과정 | 66 |
| 〈그림 3-6〉 | 뉴스 트러스트 전체 시스템 구성도 | 66 |
| 〈그림 3-7〉 | 뉴스 트러스트 개체명 사전 | 70 |
| 〈그림 3-8〉 | 뉴스 트러스트의 형태소 분석 시각화 화면 | 70 |
| 〈그림 3-9〉 | 기사 자동 분류 결과 | 72 |
| 〈그림 3-10〉 | 중복 기사 예시 | 73 |
| 〈그림 3-11〉 | 클러스터링 결과 화면 예시 | 74 |
| 〈그림 3-12〉 | 기사 제목에 사용된 단어 수 분포도 | 79 |
| 〈그림 3-13〉 | 기사 내 수치 인용 수 분포 | 80 |
| 〈그림 3-14〉 | 이용자의 요인별 저널리즘 가중치 설정 화면 | 91 |
| 〈그림 4-1〉 | 경로별 뉴스 이용 빈도 | 99 |
| 〈그림 4-2〉 | 포털 뉴스 이용 빈도(성별) | 100 |
| 〈그림 4-3〉 | 포털 뉴스 이용 빈도(연령별) | 101 |
| 〈그림 4-4〉 | 포털 뉴스 이용 목적 | 102 |
| 〈그림 4-5〉 | 포털 뉴스 이용 동기 | 103 |
| 〈그림 4-6〉 | 포털 뉴스 이용 시 주로 사용하는 기기 | 104 |
| 〈그림 4-7〉 | 알고리즘 뉴스 배열 인지 여부 | 105 |
| 〈그림 4-8〉 | 알고리즘 뉴스 배열에 관한 인식 | 107 |
| 〈그림 4-9〉 | 알고리즘 뉴스 배열의 영향력 | 108 |
| 〈그림 4-10〉 | 신뢰할 만한 뉴스 배열 주체 | 109 |
| 〈그림 4-11〉 | 뉴스 배열에서 중요하다고 생각하는 기준 | 110 |
| 〈그림 4-12〉 | 뉴스 배열의 주요 기준을 잘 수행할 수 있는 주체 | 111 |

LIST OF FIGURES

| | |
|--|-----|
| 〈그림 4-13〉 뉴스 관련 신뢰도 | 113 |
| 〈그림 4-14〉 신뢰하는 기자와 언론사 유무 | 114 |
| 〈그림 4-15〉 뉴스 신뢰도 평가 여부 | 115 |
| 〈그림 4-16〉 뉴스 신뢰도의 기계적 평가 여부 | 115 |
| 〈그림 4-17〉 뉴스 신뢰도의 기계적 평가에 찬성하는 이유 | 116 |
| 〈그림 4-18〉 뉴스 신뢰도의 기계적 평가에 반대하는 이유 | 117 |
| 〈그림 4-19〉 기계적 평가 점수를 뉴스 배열에 반영할지 여부 | 117 |
| 〈그림 4-20〉 기계적 평가 점수를 뉴스 배열에 반영하는 것을 찬성하는 이유 | 118 |
| 〈그림 4-21〉 기계적 평가 점수를 뉴스 배열에 반영하는 것을 찬성하는 이유 | 118 |
| 〈그림 4-22〉 뉴스를 접했을 때 신뢰도 평가에 영향을 미치는 요인 | 120 |
| 〈그림 4-23〉 뉴스를 접했을 때 신뢰도 평가에 영향을 미치는 요인(기사) | 121 |
| 〈그림 4-24〉 뉴스를 접했을 때 신뢰도 평가에 영향을 미치는 요인(기사) | 122 |
| 〈그림 4-25〉 뉴스를 접했을 때 신뢰도 평가에 영향을 미치는 요인(언론사) | 123 |
| 〈그림 4-26〉 뉴스를 접했을 때 신뢰도 평가에 영향을 미치는 요인(타인) | 124 |
| 〈그림 4-27〉 신뢰 지표 실험(왼쪽=신뢰 지표 포함, 오른쪽=신뢰 지표 미포함) | 126 |
| 〈그림 4-28〉 기사 작성 언론사에 관한 일반적 평가 | 127 |
| 〈그림 4-29〉 언론사의 취재에 관한 평가 | 128 |
| 〈그림 4-30〉 기자에 관한 평가 | 128 |
| 〈그림 4-31〉 언론사 만족도 및 구독 의사 평가 | 129 |
| 〈그림 4-32〉 신뢰 요인의 인지 여부 | 130 |
| 〈그림 4-33〉 신뢰 요인의 인지 여부에 따른 믿을 만한 정도 차이 | 131 |
| 〈그림 4-34〉 신뢰 요인이 뉴스 신뢰도에 미치는 영향력 | 132 |
| 〈그림 4-35〉 뉴스 배열 신뢰도 실험(점수 포함 vs. 미포함) | 133 |
| 〈그림 4-36〉 뉴스 배열 신뢰도 실험 결과 | 134 |
| 〈그림 4-37〉 뉴스 배열 신뢰도 실험(언론사명, 점수 배열방식) | 135 |
| 〈그림 4-38〉 뉴스 배열 신뢰도(언론사명, 점수 배열방식) | 136 |
| 〈그림 4-39〉 뉴스 배열 공정성(언론사명, 점수 배열방식) | 137 |
| 〈그림 4-40〉 미디어 리터러시에 관한 자기효용감 | 138 |
| 〈그림 4-41〉 뉴스 생산 과정 참여경험 | 139 |

LIST OF FIGURES

| | |
|--|-----|
| 〈그림 4-42〉 미디어 리터러시 필요성 | 139 |
| 〈그림 4-43〉 포털 뉴스 편집이 사회적으로 미치는 영향력이 유해한가 | 140 |
| 〈그림 4-44〉 포털 뉴스 편집이 사회적으로 미치는 영향력이 부정적인가 | 141 |
| 〈그림 4-45〉 포털 뉴스 편집이 사회적으로 미치는 영향력이 위협적인가 | 142 |
| 〈그림 4-46〉 포털 뉴스 배열 규제에 관한 인식 | 142 |

01

문제의 제기

영국 로이터 저널리즘 연구소(Reuters Institute for the Study of Journalism)의 <디지털 뉴스 리포트>를 인용해 한국의 언론 신뢰도가 세계 최하위라는 지적은 이제 진부할 정도다. 2016년 한국이 조사 대상에 포함된 이후 2017년부터 한국은 매년 조사 대상국 중 최하위다. 2020년도 마찬가지로 “대부분의 뉴스를 거의 항상 신뢰할 수 있다.”라는 질문에 대해 한국은 21%만이 “동의한다.”고 답해 조사 대상 40개국 평균 38%보다 크게 낮은 것으로 나타났다(박아란·이소은, 2020). 그런데, 이 결과를 조금 더 들여다보면 애매한 부분이 등장한다.

로이터 저널리즘 연구소는 각국의 언론 신뢰도를 측정하기 위해 해당 질문을 던진 후 5점 척도(전혀 동의하지 않는다 1점, 동의하지 않는다 2점, 중립은 3점, 동의한다 4점, 매우 동의한다 5점)로 응답하게 했다. 이 중 ‘동의한다’와 ‘매우 동의한다’로 응답한 비율을 합산하여 뉴스 전반에 대한 신뢰도 순위를 산출하고 있다. 그런데, 한국의 경우 ‘중립(neither agree nor disagree)’으로 응답한 비율이 45%로 40개국 평균 32%보다 유독 높았다. ‘동의한다’와 ‘매우 동의한다’의 응답률을 합산하는 방식이 아니라 5점 척도로 응답한 결과를 평균을 낼 경우 40개국의 신뢰도 순위는 달라진다. 5점 척도 기준 평균의 경우 한국은 2.80점으로 프랑스(2.71점), 칠레(2.68점), 미국(2.68점)보다 높은 36위다. 같은 조사 결과에 대해서도 해석 방식에 따라 차이가 발생한다.

뉴스 기사에 대한 신뢰는 사실 복잡적이다. 해당 기사를 얼마나 신뢰하는지는

해당 기사의 내용을 얼마나 신뢰할 수 있는지, 그 기사를 작성한 사람 혹은 언론사를 얼마나 신뢰할 수 있는지 등이 복합적으로 작용한다(민영, 2016). 이는 기사의 품질 문제와 언론의 정파성에도 연관된다. 실제 신뢰할 수 있는 정보를 전달하고 있다 하더라도 기사의 품질이 낮거나 해당 내용을 정파적으로 전달할 경우 기사에 대한 신뢰도는 낮아진다(김위근·안수찬·백영민, 2018). 기사가 전달하는 내용이 아무리 정확하더라도 바로 신뢰가 생기는 것은 아닌 것이다. 신뢰는 특정 기사에 대한 것이 아니라 그 기사를 작성한 언론사가 수용자들과 지금까지 형성해 온 관계에 좌우되기도 한다(Toff et al., 2020). 이렇듯 신뢰할 수 있는 뉴스란 무엇이며, 어떠한 기준으로 뉴스의 신뢰도를 측정할 수 있을까라는 문제는 쉽게 정의하기 어렵다.

저널리즘 연구자들은 신뢰할 수 있는 기사의 조건이라고 할 수 있는 품질 높은 기사를 측정하는 기준을 많이 연구해 왔다. 물론, 기사의 품질도 애매하고 다의적인 개념이기 때문에 학술적으로 정의하기 어렵다(박재영·안수찬·박성호, 2014). 구체적인 정의는 어렵지만 기본적으로 기사가 갖춰야 할 규범적 가치, 즉 정확성과 공정성 등을 얼마나 구현하고 있는지가 기사의 품질을 측정하는 기준으로 주로 활용되고 있다. 이를 측정하기 위한 다양한 방법론들이 제기되고 있는데, 대표적으로 미국의 저널리즘 연구기관인 PEJ(Project for Excellence in Journalism)는 2004년 기사의 품질을 평가하는 지수를 개발해 공개했다¹⁾. PEJ가 발표한 보도지수(the reporting index)는 고급기사(the highest level of reporting)를 평가하는 기준으로 활용된다.

PEJ의 보도지수는 투명 취재원(transparent source), 복합적 관점의 제시(presence of multiple viewpoints), 이해 당사자 수(number of stakeholders) 등을 종합하여 구성된다. PEJ의 보도지수는 기사 내에 투명 취재원이 4개(혹은 명) 이상 포함되고 복합적 관점이 제시되었으면, 이해 당사자가 4개(혹은 명) 이상 포함되었을 경우 고급 기사로 판단한다. 투명 취재원은 ‘기사와의 관련성이

1) <http://www.stateofthemedias.org/2005/newspapers-intro/content-analysis/>

명백히 드러난 취재원'으로서 '그가 누구인지 독자가 알 수 있는 취재원'을 말한다. 복합적 관점의 제시는 어떤 형태로든 관점이 제시되어 있었던 기사 중 '하나의 견해가 기사의 3분의 2 이상을 구성하지 못하는 경우'로, '대체로 단일 관점'은 '다른 견해가 일시적으로 언급되지만(a passing reference) 한 경우'로, '완전한 단일 관점'은 '그런 일시적인 언급조차 없는 경우'로 정의된다. 이해 당사자는 이해관계가 서로 다른 집단(different interested groups)을 의미한다.

박재영과 이완수(2010)는 PEJ의 틀을 빌려, 기사의 주제, 투명 취재원 수, 익명 취재원 수, 이해 당사자 수, 토대 정보의 신뢰성, 갈등 사안 여부, 관점 제시 양태, 인용구의 주관적 술어("말했다."가 아닌 "강조했다."처럼 발언 방향에 대한 주관적 해석), 무주체 주관적 술어 수("~로 예상된다." 등 기자의 문장에서 주체가 없는 경우) 등을 품질 높은 기사의 요소로 제시했다. 이들은 이러한 요소들을 바탕으로 토대 정보의 신뢰성이 확인된 기사, 이해 당사자가 2개 이상, 투명 취재원이 2개 이상, 인용구의 주관적 술어가 1개 이하, 무주체 주관적 술어가 1개 이하, 복합적 관점 제시(갈등 사안인 경우) 등이 이루어지면 품질 높은 기사라고 판단할 수 있다고 주장했다.

이외에도 다양한 기준을 적용하여 기사의 품질을 측정하려는 시도들이 있었는데, 이러한 시도들의 대부분은 사람이 기사를 직접 읽고 내용을 평가하여 측정하는 방식이었다. 하지만, 디지털화에 따라 뉴스 기사의 유통량이 급증하면서 네이버, 다음 등 국내 포털 사이트에 '인링크' 방식으로 등록되는 기사 수만 하루 평균 약 3만 개 이상으로 알려져 있다. 사람이 평가하기에는 너무 많은 기사다. 또한, 뉴스 유통의 중심이 구글, 네이버 등 플랫폼으로 이동하면서 플랫폼 기업 입장에서는 이렇듯 많은 뉴스를 사람이 일일이 기준에 따라 배열하기 어렵다. 이는 기사들의 품질을 기계적으로 측정하여 자동으로 배열하려는 시도로 이어졌다. 즉, 투명 취재원의 수, 이해 당사자의 수, 주관적 술어 등의 수를 기계적으로 계산해 낼 수 있다면 기사의 품질을 자동으로 측정해 그 순위에 따라 기사를 배열하는 것이다. 알고리즘에 따른 기사 배열이다. 사람이 일일이 확인해 중요한 기사와 콘텐츠를 선별하기에는 그 수가 너무 방대하다. 그렇다고 모든 콘텐츠를

나열할 수도 없고 일부만 대상으로 선별한다면 중요한 내용을 놓칠 수 있다. 이러한 문제를 해결하기 위해 적용되는 방식이 방대한 양의 기사와 콘텐츠를 빠르게 자동 선별하는 알고리즘이다

구글, 네이버, 다음 등은 100% 알고리즘에 의해 기사를 배열하고 있다고 밝히고 있다. 이들 뉴스 유통 플랫폼 기업들은 기사 배열에 있어서 사람이 간섭하지 않고 있음을 강조하면서 이용자의 선호에 따른 맞춤형 뉴스를 배열하고 있다. 여러 가지 이유가 있겠지만, 각 개인들이 보고 싶어 하는 뉴스를 위주로 전달해 더 많은 뉴스 소비를 이끌어내고 필요한 뉴스를 효율적으로 전달하겠다는 것이다. 음악, 영화, 책 등의 경우 맞춤형 콘텐츠 추천은 큰 문제가 되지 않는다. 어떤 가수의 노래, 어떤 감독의 영화만 보는 것은 취향의 문제이기 때문이다. 하지만, 뉴스라는 점에서 문제가 복잡해진다. 건강한 민주주의 공동체 사회의 유지와 발전을 위해서는 우리 사회 의견 다양성이 필요하다. 듣기 싫은 목소리도 들어야 한다. 보통 이러한 싫은 목소리들은 뉴스를 통해 전달된다. 맞춤형으로 누군가가 보고 싶어 하는 뉴스만 전달하다 보면 자신이 듣기 싫어하는 목소리를 듣지 못하게 될 가능성이 커진다. 최소한 나와 다른 의견이 있다는 사실이라도 알아야 하는데, 맞춤형 뉴스 추천 알고리즘은 그 목적상 각 개인이 그동안 관심 없었던 뉴스는 전달하지 않을 가능성이 높다. 뉴스를 많이 읽는 것도 좋지만, 다양한 의견을 듣는 것이 더욱 중요하다.

또한, 플랫폼을 통해 유통되는 기사의 양이 늘어나면서 생기는 문제 중 하나는 허위정보다. 뉴스 기사와 비슷한 형식을 취한 허위정보들이 플랫폼을 통해 유통되고 있는데, 우리가 ‘코로나19’ 상황에서 경험하고 있듯이 이러한 허위정보들은 실제 정보와 구분이 어렵다. 이로 인해 뉴스 기사에 대한 신뢰도까지 하락하는 현상이 발생하고 있다. 플랫폼 기업들은 이 문제에 대해서도 ‘자동화된 팩트체크(automated fact check)’ 기술 등을 적용해 해결하려 하고 있지만 아직 한계가 있다(오세욱, 2017). ‘팩트’는 명확히 존재하기보다 종합적으로 구성되기 때문이다(이민규, 2005). 유통되는 기사 혹은 정보의 양이 적을 때에는 사람이 그 기사 혹은 정보의 사실 여부를 검증할 수 있었다. 하지만, 현재 우리나라에서만

하루 생산되는 뉴스의 양이 수만 건을 넘는 상황에서 이를 모두 사람이 검증하는 것은 사실상 불가능하며 일부만을 선택하여 검증할 수밖에 없다. 이 때문에 현재 시점에서 한계는 있지만 자동화된 팩트체크 기술을 적용하려는 시도들이 지속적으로 이루어지고 있다. 대표적 사례가 ‘클레임버스터(ClaimBuster)’, 구글의 ‘지식 금고(knowledge vault)’, 미국 인디애나 주립대의 ‘지식 그래프(knowledge graph)’, ‘팩트마타 프로젝트(Factmata Project)’, ‘팩트마인더(FactMinder)’, ‘FiB’ 등이다.

맞춤형 기사 추천, 자동화된 팩트체크 기술 등은 모두 알고리즘에 기반한다. 알고리즘은 데이터를 처리하는 규칙이다. 어떠한 데이터가 주어졌을 때 어떻게 처리하라는 규칙의 집합으로 그 데이터를 처리하는 전략이라고 할 수 있다. 기사는 사람이 쓴다. 수없이 많은 언론사의 기자들이 쓰는 기사는 그 수만큼 형식도 다양하다. 알고리즘은 정해진 유형의 데이터를 처리할 때는 강력하지만, 뉴스의 형식은 사람마다 다를 정도로 굉장히 다양하다. 고려해야 할 요인들이 너무 많고 각 요인들은 복잡하게 뒤얽혀 있다. 같은 알고리즘을 적용하더라도 처리하는 데이터에 따라 다른 결과를 가져올 수 있다. 수많은 형식의 다양한 데이터를 처리하기 위해 알고리즘도 그만큼 복잡해질 수밖에 없다. 이렇다 보니 현재 알고리즘은 한 사람이 이해할 수준을 넘어섰다. 알고리즘이 복잡해지면서 문제가 생겨도 사람이 발견하지 못하는 경우도 자주 발생한다.

맞춤형 뉴스 추천 알고리즘이 의견 다양성 측면에서 가질 수 있는 문제, 어떤 과정을 거쳐서 결과를 제시하는지 알 수 없는 알고리즘의 복잡성과 투명성 문제 등이 있지만, 맞춤형 뉴스 추천 알고리즘의 개발과 적용은 이어질 것이다. 제시된 정보의 사실 여부를 자동으로 판단하려는 시도 또한 지속될 것이다. 효율적이기 때문이다. 한국언론진흥재단이 2016년부터 2018년까지 진행한 ‘뉴스 트러스트 위원회’는 이러한 문제의식에서 출발했다. 효율적이지만 문제도 지니고 있는 뉴스 추천 알고리즘을 보완할 수 있는 대안적 알고리즘을 개발해 보자는 것이다. “페이지뷰나 트래픽을 위한 알고리즘이 아니라, 식견 있는 공중을 위해 최적화된 알고리즘을 설계하고 구현할 수 있을까?”(Lotan, 2014, p. 118)라는 질문에서

뉴스 트러스트위원회는 출발했다. 뉴스 이용자의 선호도가 아니라 저널리즘 가치에 기반한 뉴스 추천 알고리즘을 개발해 맞춤형과는 달리 뉴스 배열에서 의견 다양성을 담아내고자 한 것이다. 저널리즘 가치에 따른 뉴스 기사 신뢰도를 계산하여 뉴스를 배열한다. 또한, 자연어 처리, 형태소 분석, 클러스터링 등 알고리즘 원천 기술을 오픈소스 방식으로 모두 공개하고 알고리즘의 기본 원칙 등도 가능한 상세하게 공개해 알고리즘의 투명성도 제고했다. 그 결과물이 ‘뉴스 트러스트 알고리즘’이다.

앞서 지적했듯이 뉴스 기사의 품질, 신뢰도 등은 쉽게 정의할 수 있는 개념이 아니다. 신뢰라는 개념 자체도 기사에 대한 신뢰인지 기사를 작성한 사람 혹은 조직에 대한 신뢰인지 등 복합적이다. ‘뉴스 트러스트 알고리즘’은 뉴스 기사 신뢰도를 계량적으로 측정하기 위한 하나의 시도였지 완전한 결과물을 제시한 것이 아니다. 이용자의 선호에 따른 맞춤형 추천이라는 것에도 정확한 답이 존재하지 않는다. 뉴스 배열 알고리즘에 있어서 100% 완성품은 사실 없다고 할 수 있다. 뉴스 트러스트 알고리즘은 저널리즘 가치 기반의 대안적 뉴스 배열 알고리즘 개발이라는 목표를 향해 가는 중간 과정이다.

이 보고서는 뉴스 트러스트 알고리즘의 개발 과정과 그 결과물에 대한 설명과 한계점을 검토하는 것을 목적으로 한다. 뉴스 트러스트 알고리즘에 대한 검토를 통해 뉴스 기사 신뢰도 측정에 있어서 문제점과 향후 어떠한 개선이 필요한지에 대해 탐색하기 위해서다. 이를 위해, 이용자들이 뉴스 기사를 읽을 때 어떠한 요인을 통해 신뢰 여부를 판단하는지에 대해서도 조사했다. 이용자들이 기사에 대한 신뢰 여부를 판단하는 데 중요하게 생각하는 요인을 밝혀 향후 이를 측정하는 데 활용할 수 있게 하기 위해서다. 뉴스 기사 신뢰도를 측정하기 위한 완벽한 정답은 없을 것이다. 다만, 이러한 작업을 통해 보다 완성도 높은 계량적인 뉴스 기사 신뢰도 측정 방안에 대해 검토해 보고자 한다.

02

계량적 기준에 따른 뉴스 기사 배열 사례 검토

1. 이론적 논의

저널리즘의 이상적인 목적은 책임 있는 권력 만들기다(Entman, 2005). 공적 책무를 수행함으로써 책임 있는 권력 만들기에 기여해야 하며 시민들에게 공적 사건에 관한 시의적절한 정보를 제공해야 한다는 것이다. 즉 시민들이 공적 사건을 평가할 수 있도록 다양한 관점을 제공해 식견 있는 시민(informed citizen) 양성에 기여해야 한다. 이에 기반해 김영욱 등(김영욱·진민정·강신규, 2014)은 국내 상황에서 저널리즘이 수행하는 기능을 [표 2-1]과 같이 정리한 바 있다.

저널리즘 품질을 구성하는 총체적 특성에 관한 연구는 크게 두 가지 관점으로 이뤄진다(김영욱 외, 2014). 먼저, 경험적 접근은 품질이 높은 것으로 평가되는 저널리즘 현장에 대한 관찰 및 저널리즘 교육자 혹은 독자들이 평가하는 저널리즘의 질적 기준을 설문과 인터뷰 등을 통해 탐색하는 방법이다. 즉 전문가, 편집자, 수용자들의 견해에 대한 관찰을 바탕으로 귀납적으로 저널리즘 품질 구성요소를 도출할 수 있다. 다음으로, 규범적 접근은 저널리즘의 사회적 기능과 역할을 전제해놓은 뒤 연역적인 방법으로 이를 수행하기 위해 필요한 덕목 요소들을 도출하는 방법을 가리킨다.

[표 2-1] 국내 상황에서 저널리즘이 수행하는 기능

| 기능 | 설명 |
|-----------------|--|
| 공표와 정보 제공 | <ul style="list-style-type: none"> ▲ 환경을 관찰해서 사회와 사회 구성원에게 중요하고 필요한 정보를 제공함 ▲ 새롭게 발생한 사건이나 새롭게 나타나거나 부각된 현상 혹은 그러한 사건과 현상을 이해하는 데 필요한 정보를 제공함 ▲ 우연히 입수한 정보가 아닌 독자와 시청자의 삶에 영향을 미칠 수 있는 환경을 체계적이고 지속적으로 관찰함 |
| 특정 문제에 대한 주목 생산 | <ul style="list-style-type: none"> ▲ 독자들이 삶의 문제를 해결하는 데 필요하거나 중요한 정보, 독자가 이 세상과 자신의 정체성을 이해하는 데 필요한 정보 혹은 독자가 관심을 갖는 정보를 선택해서 제공함. 이를 통해 특정 문제에 대한 주목을 생산함 ▲ 정보의 단순한 선택을 넘어서 특정 정보에 대해서 보도하는 양, 빈도, 방식(지면 배치 등)을 통해 보다 주목의 강도를 조정함. 이러한 과정을 통해 특정 주제가 사회적 의제로 채택될 수 있음 |
| 논의의 장 제공 | <ul style="list-style-type: none"> ▲ 일상적인 보도 행위나 논평, 인터뷰, 칼럼, 독자 편지 게재 등을 통해 특정 주제에 대한 사회적 논의의 장을 제공함 ▲ 특정 주제에 대한 의견과 견해뿐만 아니라 이를 뒷받침하는 배경 정보를 제공함 |
| 사회통합 | <ul style="list-style-type: none"> ▲ 사회 내 각 구성원과 사회 내 분화된 영역 사이를 연결해서 이들이 사회공동체 내에서 상호의존적이라는 사실을 인식시켜 줌 |
| 비판과 감시 | <ul style="list-style-type: none"> ▲ 사회에 영향을 미칠 수 있는 집단과 조직, 개인을 감시하고 비판하는 기능을 수행함 ▲ 권력 집단 혹은 권력자에 대한 감시와 비판을 위해 취재 방법과 윤리에서 높은 수준을 유지해야 함 |
| 교양 | <ul style="list-style-type: none"> ▲ 독자 자신이 살고 있는 물리적·사회적 세계를 이해하고 정체성을 성찰하는 데 필요한 지식과 태도, 심미적 경험을 하는 데 도움이 되는 지식이나 안목을 제공함 |
| 재미와 즐거움 | <ul style="list-style-type: none"> ▲ 새로운 것과 특이한 것에 흥미를 갖고 이를 추구하는 인간의 목적에 부합함 |

출처: 김영옥 외, 2014

한편, 저널리즘 품질에는 다양한 차원과 영역이 존재한다. 그만큼 저널리즘이라고 부르는 현상이 다차원적이기 때문이다. 이러한 다차원성을 반영하여 젤리저(Zelizer, 2010)는 전문직, 제도, 텍스트, 사람, 행위 집단 등 다섯 가지 유형에 따라 저널리즘 품질에 대한 논의가 달라질 수 있음을 제시하였다. 특히 저널리즘 품질을 규명하는 규범적인 접근은 저널리즘이 사회적 기능을 수행하기 위해 어떤 품질을 가져야 하는가와 관련된다. 이에 대한 대표적인 사례는 맥퀘일이 제시한 ‘미디어 퍼포먼스’다. 미디어 구조나 효과 등의 차원이 아닌 미디어가 제공하는 콘텐츠, 생산품 혹은 서비스 품질에 초점을 맞춘 것이다(McQuail, 1992). 같은 맥락에서 포트커(Horst Pöttker)는 사회 구성원들이 사회 전반에 참여할 수 있도록 기회를 제공하는 것을 저널리즘의 역할로 전제했으며 이를 수행하기 위한 저널리즘 품질을 [표 2-2]와 같이 제시했다(Pöttker, 2000; 김영옥 외, 2014 재인용).

[표 2-2] 세부차원에 따른 저널리즘 품질요소

| 차원 | 저널리즘 품질요소 | 설명 |
|---------|-------------|---|
| 보도하는 대상 | 사실성 | ▲ 정보 내용이 간주관적, 경험적 검증이 가능한 것 |
| | 완전성(중요성) | ▲ 한 사건이나 상태의 중요한 내용을 전달하는 것 |
| | 진실성 | ▲ 의혹이나 진실에 대한 판단을 제약하는 요소를 피할 수 없는 경우 이를 밝히는 것 |
| | 이질성 | ▲ 가능한 서로 다른 지적 배경이나 입장이 언급된 것 |
| 수용자·독자 | 독립성 | ▲ 저널리스트가 특정 이해를 위해 일하지 않을 것 |
| | 시의성 | ▲ 수용자의 현재 이익과 관심이 반영될 것 |
| | 독이성 | ▲ 보도의 서술 방식이 수용자에게 잘 이해되도록 할 것 |
| | 흥미성 | ▲ 적극적인 수용자를 자극하는 내용이 포함될 것 |
| 일반 윤리 | 상호성 | ▲ 해당 미디어와 내용에 대해 어떻게 피드백을 할 수 있는가를 알려주는 것 |
| | 이익 형량에서의 주의 | ▲ 저널리즘의 공표의 의무와 개인의 권리 침해 사이에서 보도 여부를 결정할 때 주의를 기울일 것 |

출처: Pöttker, 2000; 김영옥 외, 2014 재인용

이러한 저널리즘 품질에 대한 논의에 기반해 신뢰할 수 있는 뉴스는 어떻게 정의할 수 있을까? 나아가 어떠한 기준으로 뉴스의 신뢰도를 측정할 수 있을까? 기사 품질은 애매하고 다의적인 개념이기 때문에 학술적으로 정의하기 어렵다 (박재영·안수찬·박성호, 2014). 개념적 정의의 어려움에도 다양한 형태로 기사 품질을 측정하려는 연구가 진행되어 왔다. 박성호와 윤영민(2016)은 국내 연구 경향을 세 가지로 요약하였다. 첫째, 다양한 뉴스 평가 항목을 개발해 측정한 연구들로 PEJ의 기준을 활용한 연구들이 대표적이다. 박재영과 이완수(2010)는 PEJ 보도지수를 국내 언론의 환경과 맥락에 맞게 수정하면서 기사의 구성방식, 토대 정보의 신뢰성 등 20여 가지 항목으로 분석 차원을 확대했으며, ‘좋은 저널리즘 연구회’가 기획한 <기사의 품질>(김경모·박재영·배정근·이나연·이재경, 2018)은 PEJ 보도지수를 국내 현실에 맞게 수정 보완하여 한국 일간지와 해외 유력지를 비교 연구하였다.

[표 2-3] 뉴스 기사의 품질 측정에 사용된 품질 요소와 평가 방식

| 품질요소 | 평가 방식 | 연구 사례 |
|------|--|---|
| 정확성 | <ul style="list-style-type: none"> 전달 정보가 정확한지 여부를 술어를 통해 판단할 수 있음 확인된 술어 사용 여부를 기준으로 기사 신뢰도를 구분 | <ul style="list-style-type: none"> 이건호·정완규(2007) 이건호(2008) |
| 독창성 | <ul style="list-style-type: none"> 다른 언론사 기사에 없는 정보를 한 언론사가 독점하고 있는 경우 발표(사건사고를 중심으로 보도한 스트레이트 기사), 발표 기획(발표된 내용을 토대로 기자가 추가 취재를 통해 보완한 기사), 기획(기자의 독자적인 아이디어에 의해 발굴된 기사)에 따라 독창성 정도 구분 공개 정보, 재가공 정보, 단독 입수, 단독 개발에 따라 독창성 정도 구분 | <ul style="list-style-type: none"> 박재영(2006) 송용희(2005) 이건호·정완규(2008) |
| 심층성 | <ul style="list-style-type: none"> 하나의 기사에 얼마나 많은 취재원이 사용됐는지 기사가 전달하는 주제를 둘러싸고 얼마나 다양한 시각이 소개됐는지 시각을 반영하기 위해 얼마나 많은 이해관계자들이 기사에 출현했는지 | <ul style="list-style-type: none"> 박재영(2006) 이건호·정완규(2008) |
| 정확성 | <ul style="list-style-type: none"> 피동형 서술어 혹은 무주체 수동태 문장이 얼마나 많이 등장했는지 직접인용구의 주관적 술어가 얼마나 많이 등장했는지 제목에 직접인용구가 얼마나 많이 사용됐는지 | <ul style="list-style-type: none"> 박재영·이완수(2010) 이완수·박재영(2013) |
| 공정성 | <ul style="list-style-type: none"> 보편적 어휘가 아닌 은어, 전문용어를 얼마나 많이 사용했는지 정치적·이념적 편향, 가치 판단이 개입된 언어가 얼마나 많이 등장했는지 극단적인 표현, 자극적인 언어가 얼마나 많이 등장했는지 | <ul style="list-style-type: none"> 이준우·김경모(2008) 이완수·박재영(2013) |

둘째, 저널리즘의 규범적 기준을 적용해 뉴스의 전반적인 품질을 분석한 연구들이다. 주로 종이신문의 경우에는 심층성(이건호, 2008; 이건호·정완규, 2008)을 기준으로 두었고, 방송뉴스의 경우에는 객관성과 공정성(이완수·박재영, 2013)을 기준으로 적용했다. 저널리즘 규범을 기준으로 뉴스기사 품질을 분석한 연구들의 사례는 [표 2-3]과 같다.

셋째, 특정 분야의 뉴스를 선택해서 저널리즘의 규범적 가치를 적용한 품질 연구가 해당한다. 대통령 선거보도(박재영 외, 2014), 탐사보도(이민정·이건호, 2014), 방송뉴스(박재영·이완수, 2010; 이완수·박재영, 2013) 등의 연구가 대표적이다. 이들은 심층성, 공공성, 독창성 등의 가치를 기준으로 기사의 품질을 평가했다. 기사 길이 및 유형 등 형식적 측면과 함께 취재원의 질과 정보의 가치라는 기준으로 국내외 신문들을 비교분석했다. 이와 같은 연구들은 저널리즘의 규범적 가치를 기준으로 기사 품질을 측정하고 평가해왔다. 그러나 연구자가 특정한 잣대를 통해 기사를 평가하는 것은 쉽지 않다. 그 기준을 만드는 것도 쉽지 않을뿐더러, 기준을 만든 배경에 의혹을 제기할 수 있기 때문이다. 평가결과가 객관적이고 공정하지 못하다는 이유를 들며 수용하지 않는 등 오히려 평가의

취지와 관계없이 부정적인 영향을 미칠 수 있다는 우려도 존재한다(박재영·이완수, 2010).

한편, 이러한 우려와 문제점을 보완하기 위해 기사 배열에 있어서도 최근 데이터를 활용한 이용자 개인 맞춤형 서비스 개발이 증가하고 있다. 개인 맞춤형 서비스를 제공하기 위해 필수적인 기술은 추천 시스템이다. 이용자 개인의 데이터를 기계적으로 학습하여 개인별 선호도가 높을 것으로 판단되는 콘텐츠를 제공하는 방식이다. 추천 시스템은 기술적으로 협업 필터링(collaborative filtering)과 콘텐츠 기반(content-based) 방식으로 구분할 수 있다(Cremonesi, Garzotto, & Turrin, 2012; Bobadilla, Ortega, & Hemando, 2013). 협업 필터링은 이용자가 선호하는 콘텐츠를 행렬(user-item matrix)로 전환하여 특정 이용자와 유사한 콘텐츠 이용 패턴을 갖는 다른 이용자들이 많이 이용한 정보를 추천해주는 시스템이다(Cremonesi et al, 2012). 콘텐츠 기반 알고리즘은 개별 이용자가 기존에 선호한 콘텐츠와 비슷한 유형의 콘텐츠를 찾아주는 시스템이다(Bobadilla et al, 2013). 온라인 공간에서 대부분의 콘텐츠 소비방식이 이용자 개인 맞춤형으로 변화하면서 이용자들 역시 자신이 선호하는 콘텐츠 위주로 소비하는 경향이 증가하고 있다(김대원·윤영민, 2015). 이용자들은 아마존(Amazon), 넷플릭스(Netflix) 등에서 이용자의 데이터를 분석해 특정 도서·영화·음악 등을 추천하는 시스템을 적극적으로 활용하며, 국내에서도 왓챠(Watcha), 멜론(Melon) 등에서 추천하는 영화나 음악, 나아가 네이버의 에어스(AiRS), 다음카카오의 루빅스(RUBICS)에서 추천하는 뉴스를 주로 소비하고 있다. 이와 같이 추천 시스템은 전 세계적으로 온라인 이용자의 일상이 되고 있다(황용석·김기태, 2019). 이 보고서는 뉴스 기사 신뢰도를 계량적으로 측정할 수 있는 방안을 모색하는 데 목적을 두며, 다음 절에서는 뉴스 기사를 계량적인 기준으로 평가하는 국내외 사례들을 살펴보고자 한다.

2. 계량적 기준에 따른 뉴스 기사 배열

1) 해외 사례

(1) 구글

세계 최대의 검색 사이트인 구글은 검색 결과 제공을 위해 수집한 웹페이지들 중 뉴스 부분을 따로 분리해 운영하고 있다. 현재 구글은 약 70여 개 이상 국가들의 뉴스들을 제공하고 있다. 아시아만 하더라도 인도, 말레이시아, 파키스탄, 필리핀, 싱가포르, 아랍권, 아랍에미리트연합, 레바논, 사우디아라비아, 중국, 홍콩, 대만, 베트남, 일본, 한국 등 15개 국가의 언어로 된 뉴스를 제공하고 있다. 이 수많은 뉴스를 사람이 보고 가치를 판단하는 것은 사실상 불가능하다. 구글은 수집한 웹페이지 내용을 분석해 그 페이지를 참고하는 페이지가 많을수록 검색 결과 상단에 배치하는 페이지랭크 방식을 검색에 사용하고 있다. 하지만, 뉴스 부분에서는 페이지랭크 방식만을 활용하는 것이 아니라 다양한 조건들을 추가하고 있다. 예를 들어, 뉴스의 특성상 속보일 경우 아무런 내용 없이도 중요한 뉴스로 취급되기 때문에 페이지랭크 방식만으로는 중요한 뉴스를 선별하기 어렵기 때문이다.

구글이 공식적으로 밝힌 바²⁾에 따르면, 뉴스에 표시되는 내용은 명시된 경우를 제외하고 컴퓨터 알고리즘을 통해 선정된다. 이 알고리즘은 표시되는 뉴스, 이미지, 동영상 및 표시 순서를 결정한다. 구글은 뉴스 기사의 순위 책정을 위해 13가지 요인이 복합적으로 적용된 알고리즘을 활용하고 있는 것으로 알려져 있다. ‘뉴스 기사들의 순위 책정을 향상시키기 위한 시스템과 방법(Systems and methods for improving the ranking of news articles)’이라는 특허(Curtiss et al., 2013)를 통해 밝힌 13가지 요인은 언론사의 기사 생산량, 기사의 길이, 보도의 중요성, 속보성(Breaking News Score), 이용 패턴

2) Google 뉴스에서 뉴스를 선정하는 방법 Available at: <https://support.google.com/googlenews/answer/9005749?hl=ko>

(Usage Patterns), 언론사에 대한 수용자 조사(survey), 수용자 수 및 트래픽, 뉴스룸의 크기, 지국(bureaus)의 수, 실명 인용 보도의 수, 언론사의 보도 범위, 보도 뉴스의 글로벌 도달률, 글쓰기 스타일 등이다. 구글은 수집한 뉴스에 대해 13가지 요인별로 분석해 계량화된 수치를 뽑아낸다. 이 수치들을 복합적으로 적용한 알고리즘에 따라 뉴스 기사는 각각 일정한 점수를 갖게 되고 그 수치에 따라 순위를 부여받는다. 순위에 따라 이용자 화면에 뉴스 기사가 배열되고 순위 외에 분류, 시간, 화제성 등이 함께 고려된다.

이러한 순위 책정 요인은 언론사에 대한 평가와 개별 기사 자체에 대한 평가가 혼합돼 있다. 언론사 뉴스 생산량, 이용패턴, 지국 수 등 뉴스를 작성한 언론사에 대한 평가 요인들의 경우 전통이 있는 언론사의 뉴스가 새롭게 등장한 언론사의 뉴스보다 같은 주제일 경우 좋은 뉴스일 확률이 높다는 가정에 근거하고 있다. 개별 기사 자체를 평가하는 요인들의 경우, 뉴스의 길이는 심층성, 보도의 중요성은 뉴스 영향성 등과 관련이 있다. 실명 인용 보도의 수는 저명성, 글쓰기 스타일은 정확성 등과 관련이 있다.

순위 책정 알고리즘에 적용되는 요인들은 공개돼 있지만, 각 요인들이 어떤 비중으로 활용되고 있는지는 공개돼 있지 않다. 각 요인들에게 적용되고 있는 가중치는 알려진 바가 없다. 구글 특허에서는 시기에 따라 적절하게 가중치를 조정하고 있다는 언급만 있다. 매주 글로벌 이용자 분석을 통해 미세하게 변경하고 있는 구글의 검색 알고리즘(Levy, 2010)처럼 뉴스 기사 순위 책정을 위한 가중치도 매주 미세하게 변경하고 있을 것이라고 추정할 수밖에 없다. 계량화를 통한 객관적 데이터 산출로 순위를 부여하는 방식이지만, 계량화가 불가능한 공정성, 다양성 등과 관련한 부분은 포괄하지 못하고 있다.

현재 구글이 뉴스 사이트에서 제공하는 주요 뉴스의 경우는 이러한 순위 책정 알고리즘을 통해 배열하고 있지만, 계량화가 불가능한 가치 기준에 대한 보안을 위해 추천 뉴스를 별도 분류로 제공하고 있다. 구글이 제공하는 추천 뉴스는 ‘이용자가 관심 있어 할 만한 관심 분야를 토대로 자동으로 추천하는 방식이다. 이러한 뉴스 추천은 해당 이용자가 그동안 구글의 검색, 유튜브 등을 이용한

활동내역과 이용자가 직접 설정한 매체, 주제, 위치 등의 선택 내역을 활용해 이루어진다.

(2) 크레더

크레더(Credder)³⁾는 새로 나온 영화들에 대한 평가 사이트인 로튼토마토(Rotten Tomatoes)처럼 일반인들이 뉴스 기사의 신뢰도를 평가하도록 하는 미디어 스타트업이다. 저널리스트들이 평가 받고 싶은 기사를 등록하면, 크레더에 참여하고 있는 사람들이 이에 대해 ‘신뢰(trust)’, ‘비논리적(illogical)’, ‘편향됨(biased)’, ‘실수(mistake)’, ‘신뢰할 수 없는(not credible)’ 등 다섯 가지 항목에 대해 [표 2-4]와 같은 평가 이유를 선택하도록 한다. 각 항목을 선택한 후 이유를 선정하면 해당 평가가 최종적으로 기사에 대해 긍정적인지 부정적인지를 수집한다.

이러한 평가는 두 가지 방식으로 이루어지는데, 첫째, 크레더가 선정한 검증받은 저널리스트들의 평가, 둘째, 크레더의 일반 이용자들이 평가하는 방식이다. 첫째 방식의 결과는 비평적 평가(critic rating)라는 항목으로 제시되며 전체 평가 결과 중 긍정적 평가 비율이 얼마인지를 보여준다. 둘째 방식의 결과는 대중 평가(public rating)라는 항목으로 제시된다. 만약 아무도 평가를 하지 않을 경우에는 크레더가 해당 기사를 작성한 언론사 및 작성자에 대한 기존 평가 결과를 적용해 순위를 책정한다. 크레더 이용자들은 기사에 대한 평가와 해당 기사를 작성한 언론사 및 작성자에 대한 평가 정보를 함께 제시 받는다. 현재 크레더의 기사 평가는 사람이 직접 수행하는 방식이지만, 이는 이러한 결과를 기계 학습시켜 자동으로 기사를 평가하려는 목적이다.

3) <https://credder.com/>

[표 2-4] 크레더(credder)의 기사 평가 기준

| 기준 | 평가 이유 | 설명 |
|----------|--|---|
| 신뢰 | 탐사 (investigative) | ▲ 보도가 얼마나 깊이 있는지 ▲ 탐사 보도 기사를 준비하는 데 얼마나 많은 시간을 투자하는지 |
| | 훌륭한 출처 (well sourced) | ▲ 정보의 출처가 얼마나 전문적인지 ▲ 기사 본문에 근거자료를 명확하게 명시했는지 |
| | 균형(balanced) | ▲ 다양한 관점을 보도했는지 |
| | 충실한 맥락 (great context) | ▲ 사건을 둘러싼 육하원칙뿐 아니라 이면의 정보를 제공하는지 |
| 비논리적 | 추측(speculation) | ▲ 일어나지 않은 사건과 확인할 수 없는 정보를 근거로 결과를 추측 |
| | 일화적 증거 (Anecdotal Evidence) | ▲ 전체적인 맥락 고려 없이, 일회성으로 발생한 사안을 증거로 제시함 |
| | 일반화 (generalization) | ▲ 일부를 전체로 확대해서 작성한 기사 |
| | 반대편 건너뛰기 (Stacking the Deck) | ▲ 자신의 주장을 위해 반대편의 주장을 고의적으로 거부, 생략 혹은 무시함 |
| | 권위에 호소 (appeal to authority) | ▲ 권위가 있다고 인정받는 정보원의 목소리를 통해 정당화하는 기사 |
| | 허수아비 논법 (Straw Man) | ▲ 상대방의 주장을 이기기 위해 해당 사안과 상관없이 논점을 과도하여 제시함 |
| | 미끄러운 비탈길 (Slippery Slope) | ▲ 추가적이고 더 중요한 것으로 보이는 사건으로 이끄는 생각이나 행동을 제시하여 점차 진실에서 멀어지게 함 |
| 편향됨 | 감정에 호소 (Appeal to Emotion) | ▲ 결론을 지지하기 위해 논리 대신 일차원적인 감정에 호소함 |
| | 의견 (the pure opinion) | ▲ 기자의 의견에만 의존한 기사는 신빙성이 떨어짐 |
| | 인종 편향 (racial bias) | ▲ 민족, 문화, 역사, 언어 등을 이유로 특정 인물이나 집단에 관한 편견 |
| | 종교 편향 (religious bias) | ▲ 종교와 신념을 이유로 특정 인물이나 집단에 관한 편견 |
| | 재정 지원 (financial incentive) | ▲ 뉴스 주제와 관련된 인물이나 단체로부터 재정적 지원을 받은 경우 |
| | 낙시 (Hit Piece) | ▲ 객관적이고 진실해 보이는 방식으로 잘못되거나 편향된 정보를 제시해 대중의 의견을 흔들려는 시도 |
| | 국가적 편향 | ▲ 국가주의적 관점에서 개인이나 집단을 옹호 또는 비하 |
| | 젠더 편향 | ▲ 젠더 관점에서 드러나는 편향 |
| 실수 | 정치적 편향 | ▲ 정치적 입장이나 인물을 위해 왜곡함 |
| | ▲ 잘못된 이미지·영상 사용 ▲ 잘못된 용어의 사용 ▲ 사실 관계 오류 ▲ 특정 분야 과학 내용을 잘못 이해함 ▲ 연구 내용의 데이터, 분석, 주장 등을 잘못 이해함 | |
| 신뢰할 수 없음 | ▲ 풍자 ▲ 믿을 수 있는 출처 부족 ▲ 선동적이고 선정적 ▲ 성급한 일반화 | |

(3) 딥뉴스닷에이아이

딥뉴스닷에이아이(Deepnews.ai)⁴⁾는 온라인상에서 품질 높은 기사를 자동으로 찾아내는 것을 목표로 한다. 이를 위해 먼저 이용자가 기사에 대해 평가를 하도록 한다. 기계적으로 수집할 수 없는 질적 요소, 예컨대 기사의 내용이 얼마나 심층적인지, 기사가 당파적인지, 정치적으로 균형을 유지하는지, 기사가 지속적인 가치를 갖는지, 사회적 맥락과 연관성이 있는지 등을 기준으로 기사를 평가한다. 이후 평가 결과로 구축된 데이터셋을 기계가 학습하여 새로운 기사가 주어졌을 때 자동으로 기사의 품질을 평가하여 점수를 부여하고 순위를 산출하는 방식이다. 학습용 기사 평가 방식은 다음과 같다.

먼저, 이용자들은 평가할 기사를 제시한 후 6가지 항목에 대해 선택하도록 한다. 첫째, 기사 유형(Story type)으로 이용자들은 기사를 읽은 후 뉴스, 르포, 탐사, 인물·인터뷰(Feature/Profile/Interview), 사설 및 칼럼(Editorial/Opinion/Commentary), 뉴스 분석(News Analysis), 뉴스 해설(Background/Explainer), 리뷰, 해당 사항 없음(Not applicable) 중에서 하나를 선택해야 한다. 둘째, 완전성(Thoroughness)으로 뉴스 기사의 깊이가 다루는 범위의 정도를 저널리즘 관점에서 평가해 1~4점 척도로 평가한다. 1점은 낮고 4점은 가장 완전성이 높은 경우다. 셋째, 균형성과 공정성(Balance & Fairness)으로 기사가 다루는 범위에서 필요한 다양한 관점을 반영하고 있는지를 평가한다. 1~4점 척도로 1점일 경우 정파적, 4점일 경우 가장 적절히 관점 반영한 것이며, 해당 사항이 없을 경우에는 없음으로 표기한다. 넷째, 기사 수명(Lifespan)으로 해당 기사의 가치가 얼마나 오래 지속될 것인지⁵⁾를 선택한다. 다섯째, 관련성(Relevance)으로 해당 기사가 자신의 입장에서 얼마나 중요한지를 평가한다. 1~4점 척도로 1일 경우 무의미하며, 4일 경우 나와 매우 밀접한 것으로 평가하는 방식이다. 마지막은 포괄 점수(Global Score)로 흥미, 문법 등 모든 가치를

4) <https://www.deepnews.ai>

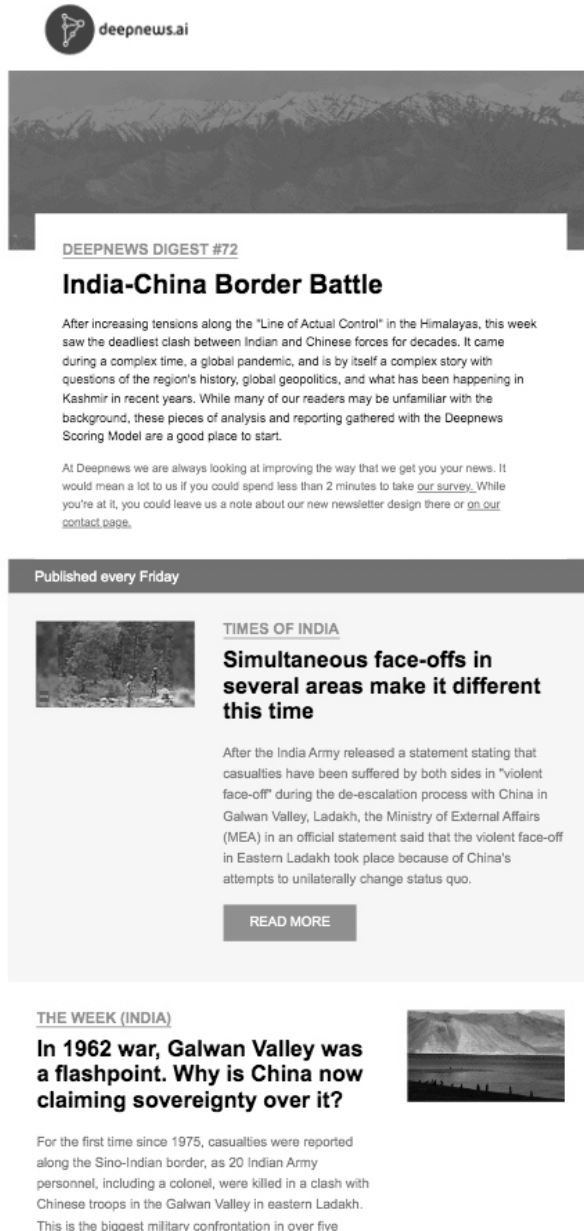
5) 1주 미만, 몇 주, 몇 달, 영원히(evergreen) 혹은 몇 년

반영하여 포괄적으로 1~4점 척도로 평가한다. 하나의 기사만 평가하는 것이 아니라 누적해서 평가할 수 있으며, 보기 싫은 기사는 건너뛰고 다른 기사를 평가할 수도 있다.

딥뉴스닷에이아이의 기사 배열은 이 결과를 학습한 후 이와 비슷한 유형의 기사를 찾아 점수를 산출하는 방식으로 이루어진다. 이들은 자동화된 뉴스 평가를 위해서는 100만 개의 학습용 뉴스 기사가 있어야만 몇천 개의 뉴스를 적절하게 평가할 수 있을 것이라고 판단하고 있다. 아직 충분한 학습 기사를 보유하지 못했기 때문에, 사람의 판단을 병행하고 있다. 현재 딥뉴스닷에이아이는 매주 50개 언론사들의 기사들을 분석해 지난 1주일간의 특정 주제를 선정한 후 자동 품질 측정 모델을 통해 수백 개 가량의 기사의 품질을 분석하고 있다. 모델에 따라 분석이 이루어지면, 해당 기사의 URL, 헤드라인, 언론사명, 기사의 단어 수, 알고리즘 점수 등의 결과가 산출되고, 이렇게 점수가 주어지면 에디터들이 수동으로 점수가 잘못된 것으로 보이거나 주제와 맞지 않는 불량 사례들을 제거한다. 예를 들어, 단어 수가 3,000개를 넘는데 점수가 1.8에 불과하거나 500 단어 미만임에도 3.9로 측정되는 경우다. 일반적인 기준에 의하면 단어 수가 많은 것이 좋은 경우가 많다. 이후 마지막으로 총괄 에디터의 최종 검토를 거쳐 [그림 2-1]과 같이 ‘딥뉴스 다이제스트(deepnews digest)’라는 제목으로 매주 기사 자동 배열 결과를 발행하고 있다.

딥뉴스 다이제스트 기사 배열 결과를 확인해 보면, 뉴욕타임스, 가디언 등 전통적인 언론사에 대한 우호가 반영되고 있음을 알 수 있다. 이는 딥뉴스 프로젝트가 기사 내 품질 요인만이 아니라 기존 언론사에 대한 이용자의 평가를 반영하고 있기 때문이다. 딥뉴스닷에이아이는 현재 자신들의 배열 알고리즘이 완성 단계가 아니며 앞으로 수많은 시행착오와 데이터 누적이 이루어져야 할 것임을 분명히 하고 있다. 이들은 이러한 결과에 대해 이용자들의 피드백을 받아 계속 수정해 나갈 계획이다.

〈그림 2-1〉 딥뉴스 다이제스트



※ 출처: <https://www.deepnews.ai/deepnewsdigest/>

2) 국내 사례

(1) 네이버

국내 뉴스 유통 시장에서 가장 높은 점유율을 차지하고 있는 네이버는 2017년 2월, 이용자의 과거 뉴스 등 콘텐츠 이용 데이터에 근거한 AiRS(AI Recommender System) 알고리즘 기반 자동 추천 뉴스를 제공하기 시작했다. 현재 네이버는 모바일 메인, PC와 모바일 뉴스홈과 섹션홈(정치, 경제, 사회, IT, 생활, 세계)에 AiRS 추천 뉴스를 제공하고 있다. AiRS 추천 뉴스에는 전체 이용자들의 기사 소비 활동을 네이버 인공지능 시스템을 통해 분석해 이용자 만족도가 높은 뉴스가 배치된다. 네이버는 딥러닝 기반의 인공지능기술인 RNN(Recurrent Neural Network)을 활용한 알고리즘으로 뉴스를 자동 편집하고 있다. 이용자들의 뉴스 소비 패턴을 예측해 데이터가 축적되지 않은 최신 뉴스의 가치를 즉각적으로 판단하고 우선 순위를 결정하는 방식이다(Park, Lee & Choi, 2017).

네이버 AiRS는 크게 협력 필터링(Collaborative Filtering) 기술과 인공지능 기반의 품질 모델(Quality Model) 기술을 적용하고 있다. 첫째, 협력 필터링 기술은 비슷한 관심 분야를 가진 사람들이 본 콘텐츠를 추천하는 기술로서, 이용자들이 콘텐츠를 소비하는 패턴을 분석한 뒤에 비슷한 관심 분야를 가진 이용자를 묶어낸다. 이 과정을 통해 이용자와 비슷한 관심사를 가진 다른 이용자들이 많이 본 콘텐츠(혹은 클러스터)를 먼저 선별해서 보여준다. 둘째, 품질 모델 기술은 인공지능기술(Neural Network)을 이용해서 정보량이 풍부하고 이용자 만족도가 높은 뉴스를 제공하는 기술이다. 여기에는 기사 요인과 이용자 피드백 요인이 활용되고 있다. 기사 요인에는 기사의 제목, 본문, 이미지, 기자 명, 작성시간 등을 활용하여 정보량이 풍부한 뉴스를 판별하고, 이용자 피드백 요인은 조회 수, 체류시간 등 다수의 이용자 소비 활동에 기반하여 만족도가 높은 뉴스를 찾는 데 활용된다. 예를 들어, 이용자 피드백 요인으로 체류시간을 예측한 후 실제 체류시간과의 차이를 통해 실제 만족도가 높은 기사를

추천하는 방식이다⁶⁾.

한편, 네이버는 2018년 3월 뉴스홈과 섹션홈 헤드라인 영역에 자동 클러스터링 기술을 적용한 기사 자동 배열을 적용하고 있다. 클러스터링은 형태소 분석을 통해 기사 본문 간 유사도를 측정하고 분류함으로써 유사한 기사끼리 묶어내는 기술이다. 네이버는 콘텐츠 제휴매체를 통해 전달받는 최신 2만 5000여 개의 기사를 10분 단위로 자동 클러스터링하여 하루에 약 1,000여 개의 이슈 클러스터를 생성하고 있다. 헤드라인 뉴스 클러스터는 최신성, 기사 길이 등 다양한 요건에 따라 배열순서가 정해지고 대표기사 역시 클러스터 랭킹 기준에 따라 추천된다. 2019년 4월부터는 'PC 뉴스홈 상단 기사(이 시각 주요 뉴스)' 및 '기존 버전 모바일 네이버 첫 화면의 기사'도 자동 추천 뉴스로 변경했다. 현재(2020년 10월 기준), 네이버 모바일 메인의 언론사 편집 뉴스를 제외한 뉴스홈(m.news.naver.com)과 섹션홈(정치, 경제, 사회, IT, 생활, 세계) 헤드라인 주요 뉴스와 하단 리스트는 모두 기계적 알고리즘을 적용해서 배열 중인 것이다.

(2) 다음카카오

다음카카오는 네이버보다 앞선 지난 2015년 5월부터 실시간 이용자 반응형 콘텐츠를 추천 시스템을 도입했다. 다음카카오는 PC 다음뉴스에 자체 개발한 'RUBICS(Realtime User Behavior-based Interactive Content recommender System, 이하 루빅스)'를 적용했는데, 루빅스는 실시간으로 이용자 반응을 분석해 콘텐츠를 추천하는 다음의 추천 시스템이다⁷⁾. 2015년 6월부터는 모바일 다음뉴스에도 확대 적용되고 있다. 루빅스는 기본적으로 뉴스 배열을 자동으로 적용하고 있지만, 기계가 완전히 수행하지 못하는 영역에 한 해 사람이 일부분 개입하고 있다는 점이 특징이다.

루빅스는 <그림 2-2>와 같이 1) 비슷한 내용의 기사들끼리 묶어주는 기술인

6) 네이버 '서비스운영원칙' <https://news.naver.com/main/ombudsman/automation.nhn>

7) <https://tech.kakao.com/2016/04/27/rubics/>

클러스터링을 적용한 후 2) 클러스터 내 기사들 중 중복되거나 어뷰징으로 판단되는 기사들을 제거하고 3) 다음카카오의 뉴스서비스 원칙에 반하는 기사들을 제외한 뒤 메인화면에 기사들을 한데 묶은 루빅스 풀을 확정한다. 이때 3번째 과정은 기계가 완전히 판단하기 어렵기 때문에 사람의 최종 검수를 거치게 된다. 이 루빅스 풀에 있는 기사들을 자동으로 이용자의 성향에 맞춰 첫 화면에 배열하는 것이다. 또한, 자동으로 식별이 불가능한 중요 기사일 경우는 사람이 직접 배열하기도 한다. 예를 들어, 대부분의 언론사가 ‘속보’로 전하는 긴급뉴스라든지, 대형사고나 재난재해, 기상특보, 스포츠중계, 경기결과처럼 실시간성이 중요한 뉴스는 전체 이용자에게 수동으로 배치하는 것이다.

〈그림 2-2〉 다음카카오 루빅스의 기사 배열 과정



※ 출처: <https://blog.kakaocorp.co.kr/412>

루빅스는 메인 뉴스의 이용자 반응을 실시간으로 측정한다. 측정 방법은 크게 전체 이용자 반응과 성·연령별 집단에 따른 반응으로 나뉜다. 다음카카오는 이를 뉴스의 중요한 속성 두 가지, 즉 사회적 트렌드와 개인적 관심사를 결합한 추천을 지향하기 때문이라고 밝히고 있다⁸⁾. 두 가지 반응은 실시간으로 기계가 학습하기에 뉴스를 많이 이용할수록 맞춤형 뉴스가 더 많이 배치되는 반면, 처음 방문하여 기사를 소비한 데이터가 적은 이용자들에게는 전체 이용자 경험을 분석해 가장 많이 본 기사가 배치된다.

8) <https://blog.kakaocorp.co.kr/412>

루빅스의 기본 알고리즘은 앞서 언급한 협력 필터링(CF) 기술이다. 개별 이용자가 소비한 뉴스와 유사한 기사를 읽은 사람들이 많이 본 기사를 추천하는 방식이다. 여기에 ‘맞춤형 멀티암드밴딧’ 알고리즘, 클릭 수 지표(CTR)-열독률 지수(DRI) 연계 등 다양한 알고리즘이 복합돼 초기 루빅스를 구성했다. 그러나 뉴스 특성상 전통적인 멀티암드밴딧 알고리즘을 적용하기 어렵다고 판단한 다음 카카오는 이동 평균 알고리즘 등을 기반으로 한 맞춤형 멀티암드밴딧 알고리즘을 개발했다. 루빅스의 맞춤형 멀티암드밴딧 방식은 성별이나 연령에 따른 이용자 집단의 뉴스 소비패턴에 근거하여 해당 집단에 속한 개인 이용자에게 뉴스를 추천하는 방식이다. 전체 이용자를 대상으로 뉴스 소비량을 계산하는 것보다 더욱 높은 정확성을 확보할 수 있다는 장점을 갖는다(박승택 외, 2017).

루빅스 자체는 이용자의 선호에 따라 기사를 자동으로 배열하는 방식이지만, 다음카카오도 이용자 요인 외에 네이버와 마찬가지로 기사 내용 요인을 활용하고 있다. 다음카카오도 클러스터링 기술을 적용하고 있는데, 클러스터 내 기사의 배열 순서를 결정하기 위해서는 이용자 요인 외에 기사의 제목, 본문, 이미지, 기자 명, 작성시간 등의 기사 요인이 필요하기 때문이다.

(3) 데이بل

국내 언론사들이 네이버, 다음카카오, 구글 등처럼 자동으로 기사를 배열하거나 추천하는 기술을 개발하기 어려운 것이 사실이다. 기사 요인 분석을 위한 자연어 처리 기술 등의 기술력과 이용자 요인 분석을 위한 데이터 등이 부족하기 때문이다. 2015년 설립된 데이블(Dable)⁹⁾은 언론사들이 자사 사이트에 추천 콘텐츠를 자동으로 배열할 수 있도록 하는 서비스를 제공하고 있다. <그림 2-3>과 같이 데이블과 제휴를 맺은 언론사는 자사 사이트 내 기사 하단이나 옆에 데이블을 활용한 자동 추천 콘텐츠 목록을 제시할 수 있다. 데이블은 자신들이 사이트에서 기사를 추천하는 것이 아니라 제휴 언론사의 기사와 데이블 서비스에 광고 게재를 요청한 광고주의 광고 콘텐츠를 혼합해 맞춤형으로 제공하고 있다.

9) <http://dable.io>

〈그림 2-3〉 조선닷컴 기사 하단 데이블의 자동 추천 콘텐츠

당신이 좋아할 만한 콘텐츠

by Dable

| | | |
|---|--|---|
|  <p>경기도 코로나로 하루 9명 사망... 신규 확진자는 244명</p> |  <p>주식투자, '이것' 따라하면 통장에 돈 쌓인다.</p> |  <p>결과도 안 나왔는데 '음성'... 집단 감염 유발 과산성모병원 고발</p> |
|  <p>[강현의 히스토리 인 팝스] [41] 밥 딜런 '구르는 돌처럼'</p> |  <p>[사설] 실세 법무차관 임명 직전 기사 폭행, 이걸 덮은 경찰</p> |  <p>문어두면 "돈" 쌓이는 주식. '이것' 시작해라.</p> |

※ 출처: <https://www.chosun.com/national/2020/12/21/DZYUELDTITVFBLS66BN6MEFVBQ/>

데이블이 이렇듯 콘텐츠를 추천하기 위해 다양한 요인을 활용하고 있는데, 뉴스 기사 추천을 위해서는 크게 7가지 방식이 활용되고 있다¹⁰⁾. 첫째, 개인 관심사 맞춤 알고리즘으로 데이블이 자체적으로 파악한 이용자의 성향을 파악해 그 이용자가 관심 있어 할 만한 콘텐츠를 자동으로 추천한다. 예를 들어, A방문자가 이전에 ‘요리’, ‘신발’ 등과 같은 목록을 검색했다면 사이트 방문 시 ‘요리’와 ‘신발’과 관련된 콘텐츠를 추천하는 것이다. 둘째, 해당 기사와 관련된 콘텐츠를 추천한다. 특정 프로야구 팀 기사라면 해당 팀의 다른 기사를 추천하는 방식이다. 셋째, 실시간 인기 콘텐츠 추천으로 해당 사이트에 그 당시 가장 인기 있는 콘텐츠를 추천한다. 넷째, 카테고리 인기 콘텐츠 추천으로 정치 기사를 클릭했다면 해당 사이트의 정치 기사 중 가장 인기 있는 콘텐츠를 추천한다. 다섯째, 성별 인기 콘텐츠 추천으로 방문자의 성별을 알고 있다면 특정 성별에서 주로 인기 있는 콘텐츠를 추천하는 방식이다. 여섯째, 연령별 인기 콘텐츠 추천으로

10) <https://support.dable.io/dable-reco-contetns/>

방문자와 비슷한 연령대의 사람들에게 인기 있는 콘텐츠를 추천한다. 일곱째, 정독 콘텐츠 알고리즘으로 해당 사이트에서 방문자가 가장 꼼꼼히 읽은 콘텐츠를 추천하는 방식이다. 방문자가 사이트에 오래토록 머무는 데 가장 크게 기여한 콘텐츠를 추천하는 것으로, 방문자 수, 본문의 길이, 체류 시간 등을 종합적으로 고려해 추천 콘텐츠를 선정한다.

3. 자동 배열 사례들의 한계

앞서 이론적 논의를 통해 저널리즘 연구들이 기사의 품질을 측정하기 위해 활용하고 있는 심층성, 공정성, 균형성 등 다양한 개념들에 대해 살펴보았다. 하지만, 자동 기사 배열 사례들을 보면 대부분은 심층성, 공정성 등 저널리즘 가치 구현보다는 이용자 맞춤형임을 강조하고 있다. 이용자 피드백 등 맞춤형 요인이 아닌 기사 요인의 경우도 기사의 길이, 작성자 명 등 기사 내용보다는 기사의 형식적 측면인 경우가 많음을 알 수 있었다. 문제의 제기를 통해 밝혔듯이 기사를 보는 관점은 사람마다 다르다. 무엇이 좋은 기사이고 신뢰할 수 있는 기사인지는 보는 사람에 따라 다르다. 예를 들어, 진보와 보수의 기사가 있다고 할 경우 두 기사를 동등한 위치에 한꺼번에 배열할 수는 없다. 어느 시점에서는 둘 중 어느 하나가 우선시 될 수밖에 없다. 그 경우 관점에 따라 그 배열은 문제가 된다. 개인마다 각기 다른 이러한 관점을 측정해서 대량의 기사 배열에 적용한다는 것은 사실상 불가능하다. 자동 기사 배열 대부분이 기사의 경중에 따른 편집이 아니라 이용자 맞춤형임을 강조하는 이유다.

이용자 요인 외에 활용되고 있는 기사의 길이, 기자 명, 제목 등 기사 요인의 경우도 문제가 된다. 기사 요인 중 측정할 수 있는 것은 제한적이기 때문이다. 예를 들어, 이론적 논의에서 기사의 품질 측정을 위해 활용된 방식 중 하나인 상반된 관점의 제시의 경우 이를 계량적으로 측정할 방법은 아직 없다고 할 수 있다. 현재 자연어 처리 기술은 기관명, 인명, 지명 등은 인식 가능하지만 기사 내용 어떤 부분이 특정 정치 성향을 띄고 있는지 등을 파악하기에는 아직 초보적 수준이기 때문이다.

구글이 뉴스 기사 순위 책정을 위해 활용하고 있는 13가지 요인이 대표적 사례다. 구글의 요인들에 따르면 전통 있는 언론사의 뉴스 기사가 새롭게 등장한 언론사의 뉴스 기사보다 같은 주제일 경우 품질이 나을 것이라는 가정에 근거하고 있다. 언론사의 작성 기사 수, 뉴스룸 크기, 지국의 수 등의 요인은 전통 있는 언론사에게 유리하게 작용한다. 기사의 길이가 길고 인용문이 많을수록

더 많은 정보를 담고 있고 보다 객관적일 것이라는 가정 등도 작용하고 있다. 나름대로 합리적 이유를 갖추고 있지만, 그대로 적용할 경우 새롭게 언론사로 등록된 ‘A’가 한국 사회에 대해 정말로 참신하면서도 갈등적 요소를 모두 포괄하고 오락성까지 구비한 파괴력 높은 기사를 작성하더라도 구글 뉴스에서 노출되기는 거의 어렵다고 볼 수 있다(오세욱, 2016).

이처럼 뉴스 기사 자동 배열 사례들은 나름대로 기준을 갖고 합리적인 방식을 적용하고 있지만, 한계를 지닐 수밖에 없다. 기사를 자동으로 배열하는 것 자체가 사실 정답이 없기 때문이다. 정답이 없는 가운데 저널리즘의 원칙을 가능한 반영하여 보다 품질이 높고 신뢰할 수 있는 기사가 우선 노출될 수 있도록 노력하는 수밖에 없다.

03

뉴스 트러스트 위원회 알고리즘 사례 분석

1. 뉴스 트러스트 위원회의 출범과 경과

1) 뉴스 트러스트 위원회 출범 배경

뉴스 유통이 디지털화되면서 온라인이 뉴스 유통의 중심 경로가 되어 왔다. 이에 따라 뉴스 소비 행태도 오프라인에서 PC를 중심으로 한 온라인으로 현재는 모바일 중심으로 변화해 왔다. 2000년 초 등장한 검색 엔진을 중심으로 한 포털은 최초 언론의 기사를 단순 중개하는 역할로 여겨졌지만 현재는 뉴스 유통의 중심이며 정보 흐름을 독과점하고 있다는 평가까지 받고 있다. 이 과정에서 포털은 언론의 역할을 수행하고 있다는 사회적 평가를 받고 있지만 스스로는 단순의 정보를 중개할 뿐이라는 입장을 고수하고 있다. 포털은 민주주의 공동체의 다양성과 균형성을 추구하고, 품질 높은 기사, 신뢰도 높은 기사의 유통을 요구받고 있지만 기사를 작성하는 언론의 책임이라고 주장하고 있다.

이 과정에서 포털의 뉴스 유통 방식도 변화하고 있다. 처음에는 언론사의 기사를 받아 포털 내 에디터를 제목을 편집하고 배열 위치를 수동으로 정하는 방식이었다. 포털의 뉴스 유통 장악력이 높아지자 이 같은 포털의 에디터 편집 모델은 포털이 사실상 언론 역할을 하고 있다는 주요 근거로 활용되었다. 이러한 지적에 따라 포털은 언론사에 포털 첫 화면의 편집을 맡기는 ‘뉴스캐스트’ 방식을

도입했지만, 언론사들 간의 트래픽 경쟁으로 인해 과도한 제목 낚시(‘충격’, ‘이럴 수가’ 등)와 트래픽을 감당 못한 언론사 홈페이지 과부하로 인해 이용자 불만이 확산되는 결과를 초래하기도 했다.

이러한 가운데 지난 2017년 10월 네이버 스포츠가 한국프로축구연맹의 청탁을 받고 축구 기사를 재배열한 사실이 드러났다. 문제가 된 기사는 2016년 10월 3일 ‘한국프로축구연맹, 누군가를 처벌할 자격이 있냐’라는 제목의 기사로 당시 네이버 뉴스 스포츠섹션 메인에 배치됐던 이 기사에 대해 한국프로축구연맹 관계자는 네이버 담당자에게 메인에서 제외해달라고 요청했고 네이버 담당자는 즉시 담당 부하 직원에게 지시해 해당 기사를 메인에서 삭제한 사실이 알려진 것이다. 이를 계기로 그동안 네이버 뉴스 편집에 대해 불만과 의구심을 드러냈던 언론사와 정치권들은 포털의 뉴스 서비스에 대해 강하게 문제를 제기했다. 네이버와 다음 등은 이러한 문제제기에 대해 뉴스 편집 공정성을 위해 전적으로 알고리즘만을 활용한 뉴스 배열을 하겠다는 방침을 밝혔고 그 결과 현재 일부 언론사의 편집판을 제외하고는 포털의 뉴스 배열은 거의 알고리즘에 의해 자동으로 배열되고 있다.

그럼에도 완전하게 문제가 해결되지는 않고 있다. 기사는 사람이 쓰기 때문이다. 수없이 많은 언론사의 기자들이 쓰는 기사는 그 수만큼 형식도 다양하다. 알고리즘은 정해진 유형의 데이터를 처리할 때는 강력하지만, 뉴스의 형식은 사람마다 달라 굉장히 다양하다. 고려해야 할 요인들이 너무 많고 각 요인들은 복잡하게 뒤얽혀 있다. 같은 알고리즘을 적용하더라도 처리하는 데이터에 따라 다른 결과를 가져올 수 있다. 수많은 형식의 다양한 데이터를 처리하기 위해 알고리즘도 그만큼 복잡해질 수밖에 없다. 이러다보니 현재 알고리즘은 한 사람이 이해할 수준을 넘어섰다. 알고리즘이 복잡해지면서 문제가 생겨도 사람이 발견하지 못하는 경우도 자주 발생한다.

현재의 포털 등이 적용하고 있는 기사 배열 알고리즘은 이용자 맞춤형임을 강조하고 있다. 여러 가지 이유가 있겠지만, 뉴스에 대한 선호가 각기 다른 이용자들이 관심 있어 하는 뉴스를 위주로 배열하겠다는 전략이다. 각 개인들이 보고

싫어 하는 뉴스를 위주로 전달해 더 많은 뉴스 소비를 이끌어내고 필요한 뉴스를 효율적으로 전달하겠다는 것이다. 음악, 영화, 책 등의 경우 소비가 늘어나는 것이 큰 문제가 되지 않는다. 맞춤형 콘텐츠 추천은 이 경우 문제가 없다. 하지만, 뉴스라는 점에서 문제가 복잡해진다. 건강한 민주주의의 공동체 유지와 발전을 위해서는 우리 사회의 다양한 의견이 필요하다. 싫은 목소리도 들어야 한다. 보통의 싫은 목소리들은 뉴스를 통해 전달된다. 맞춤형으로 누군가가 보고 싶어 하는 뉴스만 전달하다 보면 자신이 듣기 싫어하는 목소리를 듣지 못하게 될 가능성이 커진다. 최소한 나와 다른 의견이 있다는 사실이라도 알아야 하는데, 맞춤형 뉴스 추천 알고리즘은 그 목적상 각 개인이 그동안 관심 없어한 뉴스는 전달하지 않을 가능성이 높다. 뉴스를 많이 읽는 것은 물론 좋지만, 다양한 의견을 듣는 것도 중요하다.

맞춤형 뉴스 추천 알고리즘이 의견 다양성 측면에서 가질 수 있는 문제, 어떤 과정을 거쳐서 결과를 제시하는지 알 수 없는 알고리즘의 복잡성 문제 등이 있지만, 맞춤형 뉴스 추천 알고리즘의 개발과 적용은 이어질 것이다. 이용자 입장에서 뉴스 소비의 편의성만 놓고 봤을 때는 효율적이기 때문이다. 이것이 뉴스 트러스트 위원회의 출범 배경이다. 효율적이지만 문제도 지니고 있는 뉴스 추천 알고리즘을 보완할 수 있는 대안적 알고리즘을 개발해 보자는 것이다. 뉴스 이용자의 선호도가 아니라 저널리즘 가치에 기반한 뉴스 추천 알고리즘을 개발해 맞춤형과는 달리 뉴스 배열에서 의견 다양성을 담아내고자 한 것이다. 또한, 자연어 처리, 형태소 분석, 클러스터링 등 알고리즘 원천 기술을 오픈소스 방식으로 모두 공개하고 알고리즘의 기본 원칙 등도 가능한 상세하게 공개해 알고리즘의 투명성도 제고해 보고자 했다.

2) 뉴스 트러스트 위원회 논의 경과

뉴스 트러스트 위원회는 신뢰도 높은 뉴스 알고리즘 개발을 목적으로 2016년 5월부터 2018년 9월까지 운영되었다. 학계와 현장의 전문가로 구성된 위원회는 품질 높은 뉴스에 대한 사회적 합의를 끌어내기 위해 저널리즘 가치에 따라 개별 기사를 평가하고 뉴스 배열을 제공하는 알고리즘을 만들고자 했다. 3년의 기간을 거치는 동안 위원회 구성도 일부 변화했다. 뉴스 기사를 평가하려는 시도는 오래전부터 시도되었다. 미디어가 사회적 책임을 지고 있으며 공공에 이익에 이바지해야 한다는 전제에서 이루어진 품질 기준(Mcquail, 1992), 독자들이 바라본 저널리즘 품질 기준(Gladney, 1996), 나쁜 저널리즘을 기준으로 바라본 부정적 접근(Mieth, 1997), 저널리즘의 기본 원칙을 바탕으로 하는 품질 기준(Kovach & Rosenstiel, 2007/2009)이 논의되었다. 개념적인 차원을 넘어 구체적인 수준에서 품질을 측정하는 방안을 제시된 사례도 있다. 신문 품질 지수(Bogart, 1989)는 신문 편집인을 대상으로 한 설문 조사를 통해 신문 품질을 평가하는 지수를 도출했다. ‘외부 기사 대비 자사취재 기사 비율’이라든지 ‘비광고 콘텐츠 총량’과 같은 것들이 그러한 사례였다. 퓨리서치의 저널리즘 프로젝트에서 제시한 보도 지수도 대표적인 사례로 투명 취재원, 복합적 관점 제시, 이해 당사자 수를 기반으로 보도를 평가했다.

알고리즘을 만드는 것은 뉴스를 평가하는 새로운 기준을 만드는 일이기도 했다. 기존에 제시된 개념적 차원의 저널리즘 품질 기준에서 그치는 것이 아니라 기계적으로 기사를 평가할 수 있는 구체적 수준의 측정기준도 제시되어야 했다. 즉, 알고리즘에 적용할 수 있는 측정지표를 만드는 것은 명시적인 수준의 합의가 필요한 일이었다. 따라서 이러한 작업을 위한 논의는 저널리즘 영역에서 비판 없이 받아들여지는 개념에 관해 다시 한 번 생각해 보는 논쟁의 장을 제시하는 일이었다. 다양한 전문가로 구성된 위원회는 3년간 품질 높은 뉴스를 판별하는 알고리즘을 만들기 위해 다차원적인 논의를 펼쳤다. 각 연차별 주요 추진 경과는 다음과 같다. 1기 위원회의 작업은 우리가 지향하고자 하는 방향이

무엇인지를 논의하고 어떤 개념을 중심으로 뉴스를 평가할 것인지 기준을 마련하는 시간이었다. 논의 효율성을 위해 개념과 측정방식을 다루는 분과, 기사 유형 분류방식을 결정하는 분과, 논의한 내용을 기계적으로 구현하기 위한 기술적 방안을 탐색하는 분과로 나누어졌다. 이를 바탕으로 구성된 소위원회는 주요 논의사항을 결정하고 전체 뉴스 트러스트 위원과 토의를 통해 알고리즘을 만드는 과정을 진행하였다. 2기 위원회는 이용자와 위원의 기사 평가 데이터를 바탕으로 기존 논의한 내용을 확인하고 실제 알고리즘이 판별하는 결과에 대한 타당성을 검증하는 과정을 거쳤다. 3기 위원회는 기술적인 수준에서 뉴스 트러스트 위원회 알고리즘을 구현하고, 알고리즘 개발과정 전반과 소스 코드를 공개하기 위한 준비과정을 가졌다.

뉴스 트러스트 위원회는 뉴스 신뢰도에 관한 문제를 해결하는 방법의 하나로 뉴스 품질 기반의 알고리즘을 만들고자 시도했다. 일반적으로 활용하는 알고리즘은 이용자의 행동 데이터를 바탕으로 이루어지지만, 다수의 대중이 선호하는 기사가 꼭 품질 높은 기사는 아니라는 사실에서부터 시작했다. 기사의 품질을 바탕으로 판별하는 알고리즘을 만드는 과정은 기존 논의해왔던 저널리즘 평가 기준을 온라인 저널리즘 관점에서 다시 생각해 보는 기회이기도 했으며, 추상적인 차원에서 이루어졌던 주요 개념과 측정 가능한 지표의 접점을 찾아가는 과정이기도 했다. 좋은 뉴스가 무엇인지 이야기하는 것은 간단한 일이지만 위원회의 논의과정은 각자가 생각하는 좋은 뉴스에 대한 합의를 만들어가는 과정이었다. 본 장에서는 뉴스 트러스트 위원회에서 이루어졌던 다양한 논의 결과 중 논쟁적이었던 주요 사안을 중심으로 소개하고자 한다. 회의에서 이루어졌던 모든 내용을 담지는 못하지만, 알고리즘을 만드는 과정에서 발생했던 주요 논의를 통해 알고리즘이 반영해야 하는 품질 높은 저널리즘이 무엇인지에 관한 사회적 합의를 위한 시발점이 될 수 있을 것이다.

[표 3-1] 뉴스 트러스트 위원회 1기 주요 경과

| 회의 | 일자 | 주요 논의 내용 |
|-----------|---------------|---|
| 전체 회의 1차 | 2016. 05. 04. | <ul style="list-style-type: none"> • 뉴스 트러스트 사업 전반 소개 • 위원회 운영방안 |
| 전체 회의 2차 | 2016. 05. 18. | <ul style="list-style-type: none"> • 디지털 시대 저널리즘 규범적 가치에 관한 접근방식 • 알고리즘 도출 계획 |
| 전체 회의 3차 | 2016. 06. 01. | <ul style="list-style-type: none"> • '믿을 수 있는 뉴스' 개념 논의 • '믿을 수 있는 뉴스'가 담아야 할 여섯 가지 저널리즘 가치 |
| 전체 회의 4차 | 2016. 06. 14. | <ul style="list-style-type: none"> • 뉴스의 품질과 유용성 • 유용성 개념 및 평가방법 |
| 전체 회의 5차 | 2016. 06. 29. | <ul style="list-style-type: none"> • 디지털 뉴스 시장에서의 뉴스 가치평가 사례 • 저널리즘 가치 및 이용자 지표 간 접점 |
| 전체 회의 6차 | 2016. 07. 13. | <ul style="list-style-type: none"> • 기사 유형별 갖추어야 할 품질 요건 • 알고리즘 통한 저널리즘 현장 개선 가능성 |
| 전체 회의 7차 | 2016. 07. 29. | <ul style="list-style-type: none"> • 측정 가능 변수 • 분과 구성 논의 |
| 개념지수분과 1차 | 2016. 08. 12. | <ul style="list-style-type: none"> • 저널리즘 주요 개념 선정 • 뉴스 트러스트 위원회 지향점 |
| 알고리즘분과 1차 | 2016. 08. 12. | <ul style="list-style-type: none"> • 뉴스 트러스트 사업 관련 현재 기술 수준 • 주요 개념 실제 측정 방안 |
| 분류분과 1차 | 2016. 08. 12. | <ul style="list-style-type: none"> • 디지털 저널리즘의 기사 분류 범주 • 기사 유형 다양화와 분류방식의 한계 |
| 분류분과 2차 | 2016. 08. 19. | <ul style="list-style-type: none"> • 해외 언론 사례 검토 • 기사 분류방식 임시 안 논의 |
| 종합워크숍 | 2016. 08. 19. | <ul style="list-style-type: none"> • 각 분과 회의내용 종합 |
| 소위원회 구성 | 2016. 09. 06. | <ul style="list-style-type: none"> • 워크숍 쟁점 사항 정리 • 사전 및 준거 기사 진행 상황 |
| 전체 회의 8차 | 2016. 09. 08. | <ul style="list-style-type: none"> • 개별 개념 측정 방안 • 알고리즘 측정 및 기사 분류방식 |
| 소위원회 1차 | 2016. 09. 20. | <ul style="list-style-type: none"> • 개념-측정 변인 매트릭스 구성 • 개별 개념 측정 가능성 모색 |
| 소위원회 2차 | 2016. 09. 26. | <ul style="list-style-type: none"> • 개념-측정 변인 매트릭스 측정 수준 • 기존 측정항목과 미반영 측정 요인 |
| 전체 회의 9차 | 2016. 09. 29. | <ul style="list-style-type: none"> • 개념-측정 변인 매트릭스 코딩 방안 • 주요 측정 방안 검토 |
| 소위원회 3차 | 2016. 10. 13. | <ul style="list-style-type: none"> • 분석항목별 주요 측정 방안 검토 • 매트릭스 작성 방안 확인 |

| 회의 | 일자 | 주요 논의 내용 |
|-----------|---------------|--|
| 전체 회의 10차 | 2016. 10. 27. | <ul style="list-style-type: none"> • 개념-측정 변인 매트릭스 결과 보고 • 주요 측정 방안 구현 일정 |
| 소위원회 4차 | 2016. 11. 17. | <ul style="list-style-type: none"> • 언어정보 지원을 위한 사전 구축 방안 • 개념-측정 변인 매트릭스 검토 |
| 전체 회의 11차 | 2016. 11. 24. | <ul style="list-style-type: none"> • 개념-측정 변인 매트릭스 정리 • 개념별 가중치 |
| 소위원회 5차 | 2016. 12. 07. | <ul style="list-style-type: none"> • 개념-측정 변인 매트릭스 최종 검토 • 개념별 가중치 조정 방안 |
| 전체 회의 12차 | 2016. 12. 22. | <ul style="list-style-type: none"> • 기사 데이터 기초통계 분석 결과 • 매트릭스 가중치 조정 |

[표 3-2] 뉴스 트러스트 위원회 2기 주요 경과

| 회의 | 일자 | 주요 논의 내용 |
|-----------|---------------|---|
| 소위원회 6차 | 2017. 02. 15. | <ul style="list-style-type: none"> • 뉴스 트러스트 위원회 운영방안 • 뉴스 트러스트 계량 요인 추출 진행 상황 |
| 전체 회의 13차 | 2017. 02. 24. | <ul style="list-style-type: none"> • 2기 출범 • 1기 추진 경과 및 계획 보고 |
| 소위원회 7차 | 2017. 03. 09. | <ul style="list-style-type: none"> • 프로젝트 개발 현황 보고 • 프로젝트 추진계획 논의 |
| 전체 회의 14차 | 2017. 03. 30. | <ul style="list-style-type: none"> • 프로젝트 개발 현황 보고(베타 사이트 시연) • 프로젝트 수행방안 및 소위원회 구성안 |
| 소위원회 8차 | 2017. 05. 18. | <ul style="list-style-type: none"> • 계량 평가 기준 논의 • 사전 구축 논의 |
| 소위원회 9차 | 2017. 07. 05. | <ul style="list-style-type: none"> • 이용자 데이터 분석 결과 기반 알고리즘 평가 기준 • 기사 평가 데이터 확충 방안 |
| 소위원회 10차 | 2017. 07. 20. | <ul style="list-style-type: none"> • 위원회 평가 데이터 분석 결과 비교분석 • 기사 평가결과 논의 및 기타 분야 기사 평가 방안 |
| 소위원회 11차 | 2017. 08. 10. | <ul style="list-style-type: none"> • 위원회 기사 평가 데이터 분석 결과 • 기사 평가결과 및 기타 분야 기사 평가 안 논의 |
| 소위원회 12차 | 2017. 08. 22. | <ul style="list-style-type: none"> • 연예, 스포츠, 사회 기사 평가 데이터 통계분석 결과 • 의견 불일치 높은 기사 토의 |
| 소위원회 13차 | 2017. 09. 08. | <ul style="list-style-type: none"> • 정치, 경제, 사회, 국제 분야별 기사 평가 데이터 분석 • 뉴스 트러스트 기사 평가 합의안 협의 |
| 전체 회의 15차 | 2017. 10. 30. | <ul style="list-style-type: none"> • 뉴스 트러스트 계량점수 부여방안 초안 논의 |
| 전체 회의 16차 | 2017. 11. 20. | <ul style="list-style-type: none"> • 뉴스 트러스트 계량점수 부여방안 및 가중치 초안 논의 |

[표 3-3] 뉴스 트러스트 위원회 3기 주요 경과

| 회의 | 일자 | 주요 논의 내용 |
|-----------|---------------|---|
| 전체 회의 17차 | 2018. 03. 19. | <ul style="list-style-type: none"> • 뉴스 트러스트 3기 출범 • 공개 범위 및 방법에 대한 검토 |
| 전체 회의 18차 | 2018. 03. 28. | <ul style="list-style-type: none"> • 뉴스 트러스트 알고리즘 소스 코드 보고 • 공개 관련 이슈 사항 논의 |
| 전체 회의 19차 | 2018. 04. 09. | <ul style="list-style-type: none"> • 뉴스 트러스트 알고리즘 공개 추진안 • 2018년 알고리즘 개발 주요 추진계획 |
| 전체 회의 20차 | 2018. 06. 08. | <ul style="list-style-type: none"> • 뉴스 트러스트 알고리즘 3년 차 개발 계획 • 백서 작성 계획 |
| 전체 회의 21차 | 2018. 07. 06. | <ul style="list-style-type: none"> • 백서 발간 계획안 발표 • 알고리즘 고도화 사업자 선정 추진 경과 논의 |
| 전체 회의 22차 | 2018. 09. 21. | <ul style="list-style-type: none"> • 뉴스 트러스트 알고리즘 개발업체 PT • 뉴스 트러스트 알고리즘 개발을 위한 기사 평가 계획안 |

2. 뉴스 트러스트 위원회 논의 및 개발 과정

1) 뉴스 품질 평가를 위한 저널리즘 가치

(1) 저널리즘 규범 기반 알고리즘과 이용자 데이터

뉴스 품질 평가를 위해서 현업 저널리스트가 마땅히 따르고 지켜야 하는 규범이 무엇인지에 대한 논의부터 시작하는 것이 필요하다는 점에 다수 위원이 동의했다. 물론 온라인 상황에 맞는 뉴스 품질 평가 기준을 새롭게 논의하는 것이 필요하다는 점에는 쉽게 합의를 이루었고, 개념적인 차원에서 이러한 논의는 어렵지 않다고 공감했다. 하지만, 기계적으로 적용 가능한 수준에서 이러한 작업은 매우 어렵다는 사실도 모두 인지하고 있었다.

모바일로 뉴스를 소비하게 된 결과는 품질 낮은 뉴스를 소비하는 상황을 만들었다. 어떤 게 좋은 뉴스인가 규범적 가치를 논의하는 것은 위원회의 고유한 일은 아니고 이미 되어 있다. 훈련받은 사람이 평가한다면 편견이 줄어들 수 있는데 우리 결과는 기계가 할 수 있는 것이 되어야 한다. 규범 중 기계가 학습하고 적용할 수 있는 것… (전체 회의 2차, 위원 A)

품질 높은 뉴스를 판별하는 규칙을 밑바닥에서부터 만들어야 한다는 사실은 위원 모두에게 어려움을 느끼게 했고, 이용자 데이터에 의존해야 한다는 식으로 논의 흐름이 전개되었다. 전문가라고 하더라도 어떤 구체적인 측정지표를 통해 품질 평가라는 기준을 만족시켜야 할지 알 수 없었기 때문이다.

근본적 논의는 어려워진다. 학문적 논의라고 하지만 실제 온라인에서 뉴스를 이용하는 이용자들이 어떤 뉴스를 클릭하는지 어떻게 생각하는지는 알 수 없다. 클릭하기는 했지만 보면 안 된다고 생각할 수도 있다. 가능하면 언론사에서 해당 뉴스에 대한 평가 몇 가지만이라도 규범적 가치와 집어넣어서 일정 기간 관찰하면 이용자 평가가 나올 수 있다고 본다. 그걸 가지고 역으로 들어가면 어떨까? (전체 회의 2차, 위원 B)

이러한 어려움은 규칙 기반의 알고리즘을 만들어야 하는 목적이 무엇인지를 불분명하게 만들었고, 알고리즘 작성을 위해서 왜 규칙 기반의 방식을 추구해야 하는지에 대한 공감대 형성을 방해하였다. 대부분 알고리즘은 이용자 데이터 기반의 기계 학습 결과를 활용하는 추세였다. 따라서 규칙 기반 알고리즘은 과거의 방식으로 정확도와 신뢰도가 낮다는 인식이 있었고, 다수 위원은 왜 데이터를 활용하지 않는지 받아들이기 어려웠다. 최근 알고리즘과 관련된 논의는 알고리즘이 데이터가 담고 있는 편향을 그대로 반영하는 것이 아닌, 알고리즘도 적극적으로 무엇을 보여주어야 하는지에 대한 논쟁이 필요한 지점으로 바라보는 추세이다. 하지만 1기 위원회가 시작되었던 2016년도 초반에는 이러한 모든 것을 이용자 데이터 기반으로 해결하고자 하는 의견이 강했다. 당시 한 위원은 저널리즘 가치에 기반하여 평가하고 판별하는 알고리즘을 만드는 것 자체가 굉장히 엘리트적인 느낌을 준다고 말하며 부정적인 견해를 취했다.

알고리즘을 만든다는 것은 자료를 집어넣어서 결과를 보려는 기준이다. 결국에는 보고자 하는 사람이 무엇을 볼 것인가에 초점이 맞추어져 있다. 굉장히 엘리트적 느낌을 줄 수 없다. …… 모든 게 주관적이다. 기계가 보는 것을 합의할 수 있는가? 굉장히 부정적이다. (전체 회의 2차, 위원 C)

이용자 데이터 활용에 관한 논쟁은 4차 회의까지도 이어졌다. 저널리즘 가치 관점에서 알고리즘을 만드는 것에 대한 비판적 의견이 이어졌다. 반복적으로 제기된 주장의 핵심은 이용자 평가가 없다면 현실에 근거하지 않은 추상적인 논의에 머무르게 되지 않을까 하는 우려였다.

알고리즘으로 우리가 뉴스 제목과 내용 사이 차이점을 찾아낼 수 있을까. 또한, 뉴스 분류별로 평가요소가 다르기에 다시 돌아가는 이야기지만 이용자 평가를 이야기하지 않으면 좀 현실에 발을 두지 않는 논의가 될 것 같다. (전체 회의 4차, 위원 D)

규범적 논의가 이루어져야 한다는 토론은 크게 두 가지 방향으로 나뉘었다. 저널리즘 가치 기반의 평가를 이야기하는 주장은 저널리즘에서는 뉴스 소비자가 품질이 낮은 뉴스를 좋아하고 소비자에 맞추어 언론이 상업화하므로 저널리즘 품질이 악화한다고 바라봤다. 이를 전제로 받아들이는 경우 이용자 기반 평가는 문제가 있을 수 있다는 것이었다. 이용자 데이터 기반 평가를 이야기하는 주장은 이용자 데이터가 저널리즘이 추구하는 가치와 충돌한다는 전제를 가질 필요가 없다고 보았다. 클릭 수에 한정하여 바라본다면 이용자가 자극적인 뉴스만 소비한다고 볼 수 있으나 이탈률이나 체류 시간처럼 다른 지표를 다각적으로 고려하면 이용자도 자극적인 뉴스만 소비한다고 보기는 어렵다는 것이었다. 따라서 단기적인 데이터만 보면 자극적인 기사에 높은 페이지뷰가 나올 수 있지만, 장기적으로 보면 독자도 자극적인 기사만을 선호하는 것은 아니라는 견해를 보였다. 하지만 저널리즘 규범에 기반한 논의가 필요하다는 것은 위원회 초반에 이루어진 논의를 통해 확인할 수 있는 부분이다.

인터넷 기사 상당수가 기계에 맞춰 작성되고 있다. 저품질 기사가 많다. 인기검색어처럼 선정적인 것 통해 클릭을 유도할 수 있게 인간 기자들이 포털 배열 알고리즘 맞춰 작업한다. 이것을 바꾸는 것이 목적이다. 이를 위해 ‘어떤 뉴스가 바람직한 것이 아닌가’를 보여주는 게 중요하다. (전체 회의 2차, 위원 A)

이용자 데이터 반응이 제한적일 수 있다는 공감대는 5차 회의에서 생기기 시작하였다. 5차 회의에서 살펴본 실제 이용자 데이터 분석 사례는 현업에서도 높은 평가를 받는 뉴스의 특성이 전문가들이 직관적으로 판단에 의존하는 경우가 많다는 점을 보여주었다. 또한, 이용자 데이터를 수집하는 것이 나쁘지는 않지만, 데이터로 모든 것을 해결할 수 있는 것이 아니라 왜곡되는 측면 역시 많아서 실제 적용하는 게 쉽지는 않다는 점도 확인되었다. 당장은 이용자 데이터가 필요하지는 않다는 견해와 이후에도 어떤 방식의 이용자 데이터를 반영해야 할지는 신중하게 고려해야 한다는 의견이 있었다.

연예와 스포츠 뉴스가 70%를 차지하며 매체 간 전문성 차이가 거의 없다. 현재 바로 실질적 데이터 고민을 할 필요는 없지만, 일부라도 데이터를 확인할 수 있다면 저널리즘 가치와 이용자 지표들 사이에서 중간점을 찾는 데 도움이 되지 않을까 생각한다. (전체 회의 5차, 위원 D)

이용자 반응을 반영할 때 신중해야 한다. 상업적인 것 때문이다. 내가 보는 것과 추천은 다르다. 소셜미디어에서 자기 이름을 가지고 공유되는 것들은 특정 이용자 집단이라도 존중할 만한 가치가 있다. 이용자들도 자신들이 품질 높고 가치 있는 것만 공유하기 때문에, 이런 것으로 이용자 반응 확인한다면 페이지뷰나 댓글만 가지고 했을 때 자가당착 벗어날 수 있다고 생각한다. (전체 회의 5차, 위원 E)

초기 이루어진 규범 기반 알고리즘과 이용자 기반 알고리즘에 관한 논쟁은 프로젝트가 나아가야 할 방향에 대한 혼란에서 비롯했다. 몇 차례 논의를 통해 이용자 데이터가 해결해 줄 수 있는 것은 제한적이고, 온라인 저널리즘에서 필요로 하는 규범에 대해 밑바닥부터 논의할 필요가 있다는 공감대에 도달했다.

(2) 알고리즘의 접근방식 - 긍정적 접근과 부정적 접근

저널리즘 가치에 기반한 알고리즘을 만드는 것에 대한 어려움은 알고리즘이 어떤 방식을 취할 것인지에 대한 논쟁으로도 이어졌다. 프로젝트가 지향하는 목적은 뉴스 신뢰도에 관한 문제를 해결하는 방법의 하나로 뉴스 품질 기반의 알고리즘을 제시하는 것이었다. 따라서 현재 신뢰의 문제를 발생시키는 부정적인 뉴스를 걸러내는 방식으로 접근하자는 의견이 제시되었다.

좋은 뉴스인가에 대한 정의는 조금 어려울 수 있지만 안 좋은 뉴스에 대한 리스트를 생각해 보는 것은 저널리즘 품질을 떨어뜨리는 뉴스를 배제할 수 있기에 좀 더 쉬울 수 있지 않은가. 이런 게 만들어지면 포털이나 개별 언론사가 적용할 수도 있다. 출발점은 안 좋은 뉴스이다. 이런 뉴스가 관심과 트래픽 받는 거 지양해야 한다고 말할 수 있다. (전체 회의 2차, 위원 A)

긍정적인 접근보다는 부정적인 접근을 이야기했던 이유 중 하나는 현재 저널리즘 문제를 해결하는 방식이기 때문이기도 했지만 품질 높은 뉴스가 합의에 도달하기 훨씬 어렵기 때문이기도 했다. 또한, 실용적인 차원에서도 부정적인 뉴스의 기준으로 사용할 수 있는 목록을 제시하는 것이 현실적으로 가능한 결과물이라는 시각도 있었다.

사실 어떻게 믿을 수 없는 뉴스를 걸러내는지가 본질적일 수 있다. 믿을 수 있는 뉴스는 관점이 다양하다. 온라인 시대 믿을 수 없는 뉴스가 범람한다면 믿을 수 없는 뉴스 측정해서 걸러내는 것이 더 실용적이지 않을까. (전체 회의 3차, 위원 F)

좋은 뉴스라든가 이용자가 많이 보는 뉴스라든가 다 긍정적 접근이다. 정말 나쁜 뉴스라는 요소를 찾아내는 방법도 있다. 좋은 뉴스 찾으면서 나쁜 뉴스도 나오게 되어 있다. 제목과 본문이 관련성 없는 것은 조심스럽기는 하지만 공감할 수 있고 합의할 수 있는 나쁜 뉴스다. 부정적 요소가 있으면 그것도 포함을 시켜야 하지 않을까. 예를 들면 최근 혐오 표현이 기사에 등장하는 예도 있고 댓글에 무수히 많은 혐오 표현이 있는 예도 있다. 인종차별 성차별 발언 찾는 것으로 리스트만 가지고 할 수 있지 않을까. (5차 회의, 위원 G)

그렇다면 어떤 것이 긍정적 요인과 부정적인 요인이 될 수 있는지 전통적 기사 유형과 보도 원칙을 중심으로 살펴보는 과정을 거쳤다. 특히 온라인이라는 상황에서 전통적인 보도 원칙을 어떻게 적용할 수 있을지에 대한 논의가 이루어졌다. 종이신문이 기승전결 구조로 되어 있으나 온라인은 다른 구조로 되어 있으므로 온라인 이용자가 느끼는 품질은 다를 것이라고 보았다. 기사 유형별로 갖추어야 하는 원칙이 다르지만, 세부적으로 구분되는 기사 유형을 기계적으로 판별하는 것이 가능한지에 대한 문제도 제기되었다. 이러한 논의과정에서 저널리즘 신뢰 문제 해결을 위해 필요한 몇 가지 요인이 논의되었다.

실제 유통되고 있는 뉴스 사례를 바탕으로 긍정적 사례로 판별한 기사는 가치 판단을 최대한 배제한 채 사실을 전달하는 기사, 갈등 당사자들의 입장을 적절하

게 소개한 기사, 현재 사건의 과거 흐름을 정리하는 기사, 명확한 취재원을 명시한 기사, 어려운 통계 수치들을 이해하기 쉽도록 해설한 기사 같은 것이 있었다. 반면, 부정적 사례는 두세 문장의 속보 기사, 사진에 설명 한 줄만 있는 기사, 충격, 경악 등 자극적인 어휘 사용한 기사가 있었다. 그중에서도 대표적으로 논의했던 것은 익명 취재원이나 주관적 술어를 남용하는 것이었다.

익명 취재원이나 주관적 술어 남용. 그것만 해도 굉장히 뉴스 품질이 올라간다고 생각한다. 사진 하나 있어도 좋은 뉴스일 수 있다. 지금 우리가 생각하는 생산자 입장도 중요하지만, 이용자 입장도 중요하다. 중요한 것은 이용자에게 보도에 대해 얼마나 많은 이해와 정보를 제공해줬는가. 그런 관점에서 융통성 있게 활용할 필요는 있다. (전체 회의 6차, 위원 D)

익명하려고 하는데 왜 안 지켜지나. 너무 지키면 기사가 안 올라온다. 형용사, 부사나 주관적 표현은 데스크 노동이 투입된다. 데스크가 많고 꼼꼼하면 되는데 일선 기사 선에서 잘 안 지켜진다. 자기도 모르게 주관적 표현을 사용한다. 꼼꼼하게 봐주면 되는데 언론사가 권장하거나 그런 게 아니다. 안 된다고 지키라 하지만 안 지켜지는 부분이다. (전체 회의 6차, 위원 F)

알고리즘에 저널리즘에서 부정적으로 바라보는 요소를 포함함으로써 현실적으로 저널리즘 관행 개선에 도움이 될 수 있다는 시각을 가지고 있다. 하지만 현장에서 어려움을 고려하여 원칙에 어떤 유연함을 둘 것인지 고려할 필요가 있음을 이야기하고 있다. 어떤 방식의 접근을 취할 것인지에 관한 논쟁은 긍정적인 바람직한 이상향을 상정하기보다는 부정적인 것을 걸러내는 방식에서부터 출발하기로 합의하였다. 그렇지만 최종 단계에서 알고리즘이 어떤 방식으로 결과를 보여주는지는 이용자가 여러 가지 속성을 선택해서 원하는 가치를 부여하고 원하는 지표를 만드는 식으로 자체 필터링한다든지, 기사가 어떤 알고리즘에 의해 선택된 것인지 알려주고 독자가 판단할 수 있도록 하는 것이 가능한 방안이라는 합의에 도달하게 되었다.

(3) 저널리즘 가치를 반영하는 주요 개념

뉴스 트러스트 위원회에서 지향하는 저널리즘 가치가 무엇인지에 관한 논의가 이루어졌다. 기존 논의를 바탕으로 이루어진 개념적 논의는 ‘믿을 만한 뉴스’를 중심으로 해당 뉴스가 답아야 하는 저널리즘 가치를 살펴보았다. 기존 연구에 기반한 보편적인 저널리즘 가치로 정확성, 다양성, 중요성, 신뢰성, 시의성, 심층성이 언급되었고, 온라인 환경에서 고려할 저널리즘 가치로 투명성, 책무성, 독창성이 언급되었다.

여러 가지 가치를 놓고 판단한 결과 일차적으로 정확성, 다양성, 심층성, 독창성, 투명성, 책무성이 믿을 수 있는 뉴스에 관한 기준으로 언급되었다. 가치를 결정하는 과정에서 문제가 되었던 것은 상위 개념을 어떤 것으로 바라볼지에 관한 문제였다. 좋은 뉴스라든지 믿을 만한 뉴스라는 개념이 제시되었으나 두 가지 모두 이분법적으로 받아들여지기 때문에 연속선상에 있는 개념을 제시하는 것이 바람직하다는 의견이 있었다. 최종적으로는 저널리즘 가치를 담는다는 방향으로 합의가 이루어졌다.

‘믿을 수 있는’, ‘좋은’이라는 용어 모두 가치가 담겨 있다. ‘저널리즘 가치 기반 추천 뉴스’라고 지향점을 잡아서 설명할 수 있다. 어떤 뉴스에는 한 가지만 담아도 좋은 뉴스가 될 수 있고, 어떤 뉴스는 다양한 가치를 다 담아야 할 수도 있다. (전체 회의 3차, 위원 H)

다른 개념도 추가로 논의될 수 있다는 의견이 있었다. 이용자 관점에서 유용성을 고려할 수 있다는 의견도 있었고, 속보가 개별 기사로서 완성도는 떨어지지만, 유용성을 가지는 기사의 사례가 될 수 있다는 의견도 있었다. 분과별 소위원회를 구성하고 워크숍 과정을 거쳐 균형성, 다양성, 독창성, 사실성, 시의성, 속보성, 심층성, 중요성, 투명성, 책무성, 독이성, 오락성, 유용성, 접근성과 더불어 광고 윤리, 선정성, 저작권과 같은 윤리적 개념을 논의하였다. 17가지 제시된 개념 중에서 구분이 어렵거나 문제가 있는 개념을 제외하는 과정을 거쳤다. 예를 들면,

균형성과 다양성은 이론적으로는 구분되지만, 실제 측정에서는 구분되지 않는 개념이라고 보았다. 속보성과 같이 중요한 개념이기는 하지만 온라인 뉴스에서 신뢰의 문제를 발생시킨 요인을 포함해야 하는지에 대한 논의도 이루어졌다. 최종적으로는 11가지 개념을 중심으로 저널리즘 가치를 평가하고자 하였다.

개념으로부터 시작한 논의를 통해 저널리즘 가치로 11가지 개념을 제시했으나, 위원 간 논의사항은 우리가 개념에 대한 얼마나 다른 정의를 가지고 사용해왔는지에 대한 모습을 보여주었다. 예를 들면, 품질 높은 기사의 기준으로 균형성이라는 가치를 적용하는 것이 바람직한지에 대한 토론이 이루어졌다. 균형성은 저널리즘에서 오랫동안 중요한 가치로 논의됐던 것이지만, 어떤 것이 균형성이라고 말할 수 있는지에 대한 논쟁이 있었다. 기계적인 균형성이 오히려 저널리즘의 부정적인 관행으로 이어질 수 있다는 주장과 모든 영역에 기계적 균형을 적용하는 것이 바람직하지 않을 수 있다는 의견이 있었다. 균형성은 논쟁적이었으나 많은 위원이 균형성이 가지는 중요성에 공감했기에 적용 방식에 관한 논의는 계량 요인과 관계를 고려하여 결정하기로 하고 주로 고려해야 하는 저널리즘 가치에 포함하는 것으로 결정되었다.

균형성은 모든 사안보다 정치적 논쟁적 사안에서만 발생한다. 특정 영역에 관련된 기사에 만 조건에 따라 적용한다면 크게 문제 되지 않는다고 생각한다. (전체 회의 8차, 위원 I)

1975년 이후 정치 분야에서도 1:1 균형은 포기했다. 기계적 균형이 올바른가 많은 논의가 있는데 기계적으로 균형을 보는 것이 문제가 있을 수 있다. (전체 회의 8차, 위원 J)

균형성이 중요하다 주장하신 분들 계시니 일단 놔두는 게 옳다고 생각한다. 다만 이슈별로 어떤 이슈에 균형을 적용할지, 쟁점이나 논란 사항에 적용할 때 어느 한도에서 적용할지, 매체별로 차이가 필요할지는 이후에 논의하는 게 좋을 것 같다. (전체 회의 8차, 위원 B)

사례로 든 공정성에 대한 개념뿐만 아니라 다른 개념과 관련된 논쟁도 개념을 어떤 방식으로 적용할 수 있을 것인지, 상호 중복되는 부분들은 없을지에 대한

고민이 이어졌다. 사실성이나 투명성, 중요성과 심층성 같이 바라보는 관점에 따라 개념적으로 중복되는 부분들이 있다고 보았다. 따라서 완전 배타적 개념으로 바라보기에는 어려움이 있다는 점에 위원들이 공감했다. 하지만 뉴스 이용자에게 품질 높은 뉴스를 판별하기 위해 다양한 개념을 고려했다고 설명하는 것도 프로젝트의 목적이기 때문에 실제 측정과 관련하여 각 개념의 활용 가능성을 다음에 정리하는 식으로 합의가 이루어졌다.

2) 계량 요인과 저널리즘 가치의 연결

(1) 활용 가능한 측정 요인의 정확성

기사 품질을 판별하는 저널리즘 가치를 무엇으로 바라볼 것인지에 관한 논의 결과를 바탕으로, 이후 이루어진 작업은 뉴스 기사에서 기계적으로 측정할 수 있는 요인을 논의된 저널리즘 가치와 연결하는 과정이었다. 개념에 관한 논의과정을 통해 11가지 주요 저널리즘 가치를 선정했으나, 기계적으로 측정 가능한 요인이 추상적 수준의 저널리즘 가치와 어떤 방식으로 연결될 수 있을지는 위원 모두 어려움을 느꼈다. 지속해서 이루어진 회의에서 향후 어떤 방향으로 논의가 이루어져서 실제 가능한 결과물을 만들어낼 수 있을지에 대한 불확실성을 보였다.

개념부터 나와서 연역적이고 추상적이다. 그다음은 밑에서부터 올라오는 것이어야 한다. 측정 가능한 요인을 펼쳐놓고 개념과 만나게 하는 것이다. 여기까지 왔으니 워크숍이 필요하다. 잠깐 해서 해결될 건 아니다. 측정 가능한 요인을 들고 와서 다 쏟아놓고 겹쳐는 것, 차이도 있고 가치 기준 어디 끼워 맞출 수 있을까. 새로운 용어 개념 만들어야 한다. 우리가 필요하고 할 수 있는 게 뭔가 나오지 않을까. (전체 회의 3차, 위원 H)

개념과 측정 요인을 연결하기 위해 기술적으로 우선 적용 가능한 요인으로 어떤 것들이 있는지를 확인하고, 이를 연결하는 방안에 대한 방식으로 논의가 이루어졌다. 정확성에 대한 어려움은 기계 학습 방식을 적용하면 해결할 수 있지

않은지 다수의 의원이 이야기했으나, 실제 기계 학습이 모든 것을 해결해 줄 방안이 아니라는 이야기가 이루어졌다.

컴퓨터가 구분해줄 수 있는가? 저널리즘 좋은 뉴스 기준과 알고리즘 판별이랑 너무 다르다. 이 차이를 어떻게 줄이는가. (전체 회의 4차, 위원 D).

알고리즘을 가정해서 논의하면 너무 제한적이다. 왜곡될 수 있는 부분이 있다. 그렇게 생각하면 판별할 기회를 놓칠 수 있는 위험이 있다. 좋은 뉴스 정의를 사람이 보기에 타당하게 자세히 하면 그걸 알고리즘이 어떻게 찾을까를 고민하며 가야 한다. 알고리즘 할 수 있는 것이 제한적이다. 예를 들면, 낚시성 제목. 간단한 기계 학습 알고리즘을 해봤는데 정확도가 70%를 넘지 못한다. 현실적으로 이걸 신뢰성 점수로 줄 수 있는 수준은 아니다. 꽤 유의한 지표를 찾아서 점수를 붙여 내주는 게 지금 기술에서 가장 현실적인 방법이다. (전체 회의 4차, 위원 K)

언론사에서 알고리즘을 활용하여 기사를 측정하고 판별하는지 다양한 실제 사례를 중심으로 논의가 이루어졌다. 해커뉴스(HackerNews)의 기사 평가 알고리즘, 구글의 속보 기사 평가, 제목-본문 일치도 평가, 콘텐츠 소비율 평가, 카드뉴스 상호작용성 평가, 시계열에 따른 노출과 상호작용 평가와 같은 현장의 콘텐츠 기반 측정 방안을 살펴보았다. 이를 통해 확인한 결과는 기존 연구에서 밝혀진 명확한 근거가 있는 것이 아니라 전문가가 직관과 감각으로 결과를 판단한다는 점이었다.

특정 밝혀낸 게 전문가 직관과 감각으로 한 판단이지 사회과학자들이 그것 때문에 사이트에 머물렀다고 구체적으로 손에 잡히는 자료를 가지고 말한 건 아니지 않은가? (전체 회의 5차, 위원 H)

측정치가 뉴스 저널리즘 가치와 어느 정도 맞아떨어지는지는 알기 어렵다. 내부 사람들이 경험과 직관을 바탕으로 측정치와 비교하며 진행한다. (전체 회의 5차, 위원 L)

[표 3-4] 측정 가능한 요인 초안

| 측정 변인 | 측정 내용 |
|-----------------|--|
| 기사 길이 | 더 많은 정보를 담은 기사라고 판단 |
| 인용 수 | 기사 내 인용문이 많을수록 객관성 및 다양성 등이 높으며, 기사 사건이 개입하지 않은 것으로 판단 |
| 정보원 수 | 기사 내 정보원 수가 많을수록 기사 정확성 및 신빙성 높음 |
| 기자 명 여부 | 언론사 소속 기자가 실명으로 쓴 경우 믿을 수 있음 |
| 맞춤법 및 비문 | 맞춤법 준수할수록 좋은 기사로 판단 |
| 제목과 본문 일치도 | 일치도가 낮을수록 과장성이 높다고 판단 |
| 문장당 평균 형용사·부사 수 | 높을수록 정보전달 명료성 떨어진다고 판단 |
| 문장당 평균 접속사 수 | 많을수록 간결성 떨어진다고 판단 |
| 문장당 평균 단어 수 | 단어가 많을수록 가독성 떨어진다고 판단 |
| 링크 및 멀티미디어 수 | 많을수록 정보성 높다고 판단 |
| [단독], [속보] 컷 | 단독 및 속보 파악 |

모든 위원이 기계 학습이나 알고리즘 전문가가 아니므로 기술이 해결할 수 있는 부분에 관한 오해가 있었으나, 아무런 기본이 없는 상태에서 모든 것을 처음부터 논의하여 재검토하고 결정하는 과정이 위원회 내부에서 필요하다는 사실에는 합의에 도달했다. 따라서 이후 이어진 회의는 현재 기술적으로 측정 가능한 요인을 개념적인 내용과 연결하는 초안을 만들었다.

측정 가능한 요인 초안으로 제시된 내용은 [표 3-4]에서 보는 것과 같다. 해당 요인은 기사에 대한 전처리가 완료된 후 기초적인 수준에서 기계적으로 추출할 수 있는 요인으로 구성되었다. 기사 길이, 인용 수, 정보원 수, 기자 명 유무, 맞춤법 및 비문, 제목-본문 일치도, 문장당 품사 수, 링크 및 멀티미디어 수, 제목 컷 내용과 같은 것이 기초적으로 추출할 수 있는 요인으로 제시되었다.

기초적인 측정 요인이 제시된 이후에 정확도에 대한 우려는 더욱 커졌다. 위원들은 추상적인 저널리즘 가치와 구체적인 측정 요인이 적절하게 연결될 것 같지 않다고 보았고, 정확도를 높이기 위해서는 더 많은 요인을 고려할 필요가 있다는 의견을 밝혔다.

추상화 수준들이 다른데 개념은 높은 부분, 컴퓨터가 계량화할 수 있는 것들은 수준이 낮다. 그대로 뽑아낼 수 있는 것 주셨는데 그들 간 대응이 정확히 그려지기는 어려울 것 같다. 조금 더 고민이 필요하다. 알고리즘 정확도를 높일 수 있는 부분들이 있고, 키워드를 이용한 부분이 들어가 있지 않는데 고민해봐야 할 것. (전체 회의 7차, 위원 K)

변인들을 이것보다 더 많이 제시해줄 수 있는가? 제한적이다. 개념적인 것과 연결하는 게 어렵다. 개연성이 떨어진다. 개발할 수 있는 변수 종류를 10배는 해줄 수 있는지. 이런 것을 할 수 있다고 하면 반대로 해서 연결하는 게 현실적일 수 있다. 우리가 한다고 하는데 못하면 정확도가 떨어진다. 중요한 게 결과치이다. 아무리 좋아도 정확도가 떨어지면 신뢰도가 떨어진다. 개념적 논의를 줄여서 가능하면 기준을 적게 적용하는 게 정확도가 높아질 수도 있다. (전체 회의 7차, 위원 C)

기사 전처리기가 끝난 후 다양한 분석을 통해 측정 가능한 요인을 추가하는 방향으로 프로젝트를 진행하기로 하였으나, 정확도에 대한 우려는 기사 내부 요인 이외에 다른 측정 요인을 알고리즘에 반영하는 것에 대한 논의로까지 이어졌다.

(2) 기사 외적 측정 요인 반영

기사 내부 요인만으로 기사의 품질을 평가하는 것이 어렵고 정확도가 높지 않을 수 있다는 우려로 기사 외적 요인을 결합하여 평가하는 방안에 대한 논의도 이루어졌다. 기사 외적 요인으로 언급된 것 중 대표적인 사례는 언론사에 관한 평가였다.

트러스트 활동 목표가 기사 품질과 신뢰도를 높이기 위함이다. 그것을 담아낼 수 있는 알고리즘을 만드는 것이 중요하다. 정량적 계량화 방법. 사람들이 공통으로 품질 높은 뉴스라고 평가하는 것이 정량화되면 반영해보는 게 어떤가. 구글 뉴스는 기사를 얼마나 많이 생산하는지, 편집국 규모, 언론사 신뢰도를 하나의 요인으로 활용한다. 우리도 이런 것을 고려해보면 어떤가. 언론사 역사. 100년과 6개월 언론사의 신뢰도를 다를 수 있다. 언론사 수상은 기자와 독자가 좋은 기사라고 평가한 결과이다. 그런 걸 많이 생산한 기자라든지 어떤 매체가 받았다는 결과... (전체 회의 7차, 위원 A)

실제 구글 뉴스에서 활용하는 알고리즘에 언론사와 관련된 요인이 있다는 것이 알려져 있으나 외적 요인을 포함했을 때 공정성을 어느 정도 담보할 수 있는지 반론이 제기되었다.

포함하자는 것 충분히 이해한다. 우리가 하는 것은 우선 뉴스 내용을 가지고 하는 작업이다. 이것이 확장해서 언론사를 포함하는 것, 중요하지 않은 게 아니라 다음 단계이고 별도 영역이라고 생각한다. (전체 회의 7차, 위원 H)

다양하게 제시된 의견 중 정정 보도와 같은 것을 알고리즘 평가에 반영하자는 의견도 제시되었다. 정정 보도에서 일반적으로 사과표현을 잘 하지 않는 관행이 있다는 의견이 제시되었고, 정정 보도의 진정성을 평가하여 반영하자는 안이 논의되기도 했다. 하지만 이러한 기사 외적 요인도 평가에 포함하는 것과 관련해 회의적인 의견이 제기되었다. 언론사 보도 관행에서 정정 보도를 자체적으로 하는 경우는 많지 않고 언론중재위원회 권고 사항인 경우가 많다는 것이다. 또한, 개별 뉴스 기사 평가에서 확인할 수 있는 것이 아니라 언론사의 신뢰도와 관련된 지표이기 때문에 언론사 요인으로 연결될 수 있다는 것이다.

취재원에 대한 평가를 반영하자는 의견도 있었다. 보도 내 개체명을 분류하고 해당 개체명이 누구를 지칭하는지 기술적으로 판별하기 위해서는 기사에서 개체명을 정확하게 분류해내는 기술과 함께 개체명에 관한 정보가 필요했다. 전문가가 아닌 여론을 담는다는 목적으로 익명의 온라인 아이디나 보통 시민 명칭을 사용하는 것이 문제가 된다는 의견도 있었다. 하지만 이러한 내용 역시 기사 품질 평가를 위한 알고리즘에 적용하기는 어려운 부분이었다. 기술적인 정확도를 제외하고서라도 해당 개체명이 시민인지 전문가인지 기계적으로 판단하는 것이 높은 난이도를 가지는 작업이었기 때문이다. 다양한 요인이 고려되었으나 합의 과정을 거쳐 가능한 측정 개념을 늘어놓고 어떤 개념과 연결할 수 있을지 확인하는 작업을 진행했다.

(3) 측정항목과 가치 개념의 연결작업

위원회에서 논의한 11가지 개념과 개발팀에서 향후 개발과정을 거쳐도 불가능하다고 의견을 제시한 몇 가지 측정 요인을 제외하고 개념과 측정 방안을 모두 포함하는 매트릭스를 만들었다. 즉, 기술적으로 현재 상황에서 가능한 것과 이후 추출 가능할 것으로 예측하는 요인을 모두 포함했다. 측정 요인 후보군은 [표 3-5]와 같았다. 이러한 요인은 저널리즘 가치를 논의하는 과정에서 제시된 부정적인 저널리즘 관행에 관한 내용을 포함했다. 이와 같은 측정 요인을 바탕으로 기존 위원회 논의과정을 통해 이야기했던 저널리즘 가치 요인과 연결하는 방안을 이야기하기 시작했다.

[표 3-5] 위원회에서 논의한 측정 요인 안

| 측정 요인 |
|---|
| [기획]이라는 컷 사용이 있는 경우 |
| [단독]이라는 컷 사용이 있는 경우 |
| 검색 키워드를 많이 사용할수록 |
| 공유수가 높은 기사일수록 |
| 관련 기사 링크 제목이 선정적일수록 |
| 광고와 기사가 명확하게 구분되는 기사일수록 |
| 기사 내 공적 인물의 수가 많을수록 |
| 기사 내 취재장소가 명기되었을 경우 |
| 기사 안에 자극적인 내용이 많을수록 |
| 기사당 문장 수가 많을수록 |
| 기사에 등장한 개체명(인물, 기관, 장소)이 많을수록 |
| 기사의 길이가 길수록 |
| 기자의 실명이 적힌 기사인 경우 |
| 기자회견, 간담회, 인터뷰 등 기사의 형식을 파악할 수 있는 기사일수록 |
| 기존의 다른 기사와 중복이 심할수록 |
| 다른 기사와는 구분되는 고유정보가 기사 내에 많이 포함될수록 |
| 댓글 수가 많은 기사일수록 |
| 등장한 모든 정보원의 수가 많을수록 |
| 링크된 관련 기사와 원 기사 간의 실질적 관련성이 적을수록 |
| 맞춤법이 철저히 준수될수록 |

| 측정 요인 |
|--|
| 무주체 술어(알려졌다, 전해졌다) 사용이 많을수록 |
| 문장당 평균 단어 수가 많을수록 |
| 문장당 평균 부사 수가 많을수록 |
| 문장당 평균 접속사 수가 많을수록 |
| 보도자료의 문장과 일치도가 높은 기사일수록 |
| 비속어 사용이 많을수록 |
| 사용한 정보원의 차별성이 뛰어날수록 |
| 사진 등 이미지 자료 수가 많을수록 |
| 상반된 의견의 정보원 인용 비율이 비슷할수록 |
| 선정적 표현 사용이 많을수록 |
| 수치 인용이 많을수록 |
| 실명 인용 수가 많을수록 |
| 실명 정보원의 수가 많을수록 |
| 앗, 이럴 수가, 충격 등 감정 노출용 단어를 제목에 많이 포함한 기사일수록 |
| 익명 인용 수가 많을수록 |
| 익명 정보원의 수가 많을수록 |
| 인식된 개체명이 중요할수록 |
| 인용문의 길이가 길수록 |
| 인용문의 수가 많을수록 |
| 일반 시민 발언 인용이 많을수록 |
| 일정 중복도 이상 기사 중 가장 빠른 기사일 경우 |
| 쟁점 사안에 대한 긍정, 부정 발언 비율이 비슷할수록 |
| 정보원 전문성이 높을수록 |
| 정보도시 사과나 유감 표현의 진정성이 분명하게 드러날수록 |
| 제목과 본문 내용이 일치할수록 |
| 제목에 부사를 많이 사용한 기사일수록 |
| 제목에서 따옴표, 느낌표, 물음표 등을 많이 사용한 기사일수록 |
| 조희 수가 높은 기사일수록 |
| 주관적 술어 사용이 많을수록 |
| 첨부된 표가 많을수록 |
| 카드뉴스 형식의 기사인 경우 |
| 타사의 추종(뒤늦게 따라가는) 보도를 많이 초래한 기사일수록 |
| 하이퍼링크 수가 많을수록 |
| 혐오적 표현 사용이 많을수록 |
| 활용한 멀티미디어 수가 많을수록 |

측정 요인은 일정한 방향성을 가진 명제 형태로 제시되었다. ‘기사 길이가 길수록’이라는 측정 요인이 각 개념과 가지는 관계를 평가하는 방식이었다. 논쟁이 되는 부분은 각 측정 요인이 개념과 어느 정도 관련성을 가지는지 코딩하는 방식에 관한 것이었다. 기본적인 코딩 원칙을 정하는 과정에서도 위원 간 시각 차이가 존재했다. 측정 단위와 방향성에 관한 부분에서 이견이 있었다.

매트릭스 측정항목과 변인 사이 관계성을 말하는지 정적 부정 관계를 이야기하는지 혼동 …… 매트릭스 측정에서 긍정 부정을 나누기 힘든 것도 있을 수 있지만, 현재 방향 표시할 수 있는 시스템으로 정하고 방향성 나타낼 수 있는 스케일로 정하는 것이 좋을 것 같다. (소위원회 2차, 위원 H)

부적 코딩을 해야 한다면 5점 척도일 때 -2에서 2까지 하는 것이 좋다고 생각한다. 스케일이 너무 넓으면 문제가 생기고 바이너리로 코딩하면 변량을 주기 힘들니 5점 정도가 적당하다고 본다. (소위원회 2차, 위원 D)

방향성을 주는 것에 부정적인 의견도 있었다. 못하는 기사에 감점하기보다는 잘하는 기사에게 가점을 주자는 것이었다.

매트릭스 측정 스케일에서 정적 부정에 해당하지 않는 개념도 있을 것 같다. 오히려 별점으로 0부터 5까지 측정하는 것들이 정착되었기 때문에 긍정 부정보다는 별점으로 하는 것이 좋다고 생각한다. 잘하는 기사에게 가점을 주는 쪽으로 가야 많은 문제 풀 수 있다. …… 예를 들면, 익명은 부정적, 실명을 긍정적으로 보는 것은 반대한다. 실명에 점수를 주어야 한다고 생각한다. (소위원회 2차, 위원 E)

논란의 여지는 있었으나 우선은 대략적인 개념과 측정 요인 연결을 위해 일괄적으로 코딩을 진행하기로 합의하였다. 긍정적인 각 측정 방안이 개념을 측정하는지 얼마나 관련성을 가지는지에 따라 -2점부터 2점까지 점수를 부여하는 과정을 거쳤다. 예를 들면, 기사 길이는 심층성과 밀접한 관계가 있기에 2점으로 볼 수 있지만, 균형성과는 관련성이 없으면 0점을 부여하는 식이었다.

소위원회에 참여한 위원들이 요인과 개념 간 매트릭스에 평가한 결과 평균값 중 편차가 큰 요인에 대해 전체위원회 논의를 다시 거치는 과정을 가졌다. 의견 차이가 컸던 부분은 익명 정보원과 문장 품사와 관련된 내용이었다. 익명 정보원에 대한 편차가 크기에 익명 정보원이 많다고 해도 좋게 볼 수 있는 부분이 있는지 논의했다.

〈그림 3-1〉 위원회에서 논의한 측정 요인 안

| 개념 | 측정요인 | 평균값 |
|-----|--|-------|
| 투명성 | 광고와 기사가 명확하게 구분되는 기사일수록 | 1.25 |
| 투명성 | 기사 내 공적 인물의 수가 많을수록 | 1.25 |
| 투명성 | 기사 내 취재장소가 명기되었을 경우 | 1.58 |
| 투명성 | 기사 안에 자극적인 내용이 많을수록? | -1.00 |
| 투명성 | 기사에 등장한 개체명(인물, 기관, 장소)이 많을수록 | 1.08 |
| 투명성 | 기자의 실명이 적힌 기사인 경우 | 1.33 |
| 투명성 | 기자회견, 간담회, 인터뷰 등 기사의 형식을 파악할 수 있는 기사일수록 | 1.25 |
| 투명성 | 다른 기사와는 구분되는 고유정보가 기사 내에 많이 포함될수록 | 1.00 |
| 투명성 | 등장한 모든 정보원의 수가 많을수록 | 1.33 |
| 투명성 | 링크된 관련기사와 원 기사간의 실질적 관련성이 적을수록 | -1.08 |
| 투명성 | 무주체 술어(알려졌다, 전해졌다) 사용이 많을수록 | -1.50 |
| 투명성 | 수치 인용이 많을수록 | 1.42 |
| 투명성 | 실명 인용 수가 많을수록 | 1.50 |
| 투명성 | 실명정보원의 수가 많을수록 | 1.67 |
| 투명성 | 앗, 이럴수가, 충격 등 감정노출용 단어를 제목에 많이 포함한 기사일수록 | -1.08 |
| 투명성 | 익명 인용 수가 많을수록 | -1.58 |
| 투명성 | 익명 정보원의 수가 많을수록 | -1.58 |
| 투명성 | 인용문의 수가 많을수록 | 1.17 |
| 투명성 | 정보원 전문성이 높을수록 | 1.08 |
| 투명성 | 정정보도시 사과나 유감 표현의 진정성이 분명하게 드러날수록 | 1.25 |
| 투명성 | 주관적 술어 사용이 많을수록 | -1.33 |
| 투명성 | 첨부된 표가 많을수록 | 1.08 |
| 투명성 | 하이퍼링크 수가 많을수록 | 1.25 |

수치상으로 편차는 있었으나 ‘~관계자’, ‘~ 측’ 같은 용어를 어쩔 수 없는 경우 아니면 될 수 있으면 하지 말아야 한다는 의견이 다수였다. 물론 어쩔 수 없이 익명을 사용할 수밖에 없는 현장의 사정도 존재했으나, 위원회에서는 원칙상 익명 취재원이 많은 것이 좋은 것은 아니라는 점에 합의했다.

품사의 경우는 현장에서 사용하는 원칙과 일반 위원이 느끼는 품사 활용에서 차이가 있었다. 독자 관점에서 기사를 읽을 때 형용사나 부사가 많으면 읽기 편하다는 의견이 있었으나, 현장에서는 접속사를 걸러내고 꼭 필요하지 않으면 쓰지 않는다고 밝혔다. 간략하게 설명해야 하는데 중언부언하면 친절하지 않다고 생각하며 정보가 유익적이지 않고 나열이 되기 때문에 접속사가 빠지는 것을 좋은 문장이라고 보았다.

논의를 통해 상반된 결과가 나온 것들에 대해 합의를 통해 결정하고 평균값이 절댓값 1 이상 나온 요인을 중심으로 개념의 측정 요인을 결정하였다. 물론 기술적으로 모든 측정 요인을 사용할 수 있던 것은 아니다. 기술적으로 적용 가능한 부분부터 단계적으로 시스템을 만들어나갔다.

3) 알고리즘 평가에 대한 합의 과정

(1) 점수 산정 기준

저널리즘 가치를 측정하기 위한 다양한 개념과 측정 요인이 가지는 관계를 고려하는 과정을 거친 이후 필요한 작업은 각 요인에 따른 점수 부여에 관한 문제였다. 즉, 측정 요인에 따라 어떤 기준으로 가점을 주거나 감점을 줄 것인지 고민하는 과정을 거쳤다. 가장 어려운 부분은 항목별 점수를 산정하는 기준을 어떻게 정할 수 있는지에 관한 내용이었다. 기준을 정하는 것이 논리적인 근거나 판단 기준 없이 자의적일 수 있기 때문이다. 따라서 계량 요인 중 단순 추출을 통해 확인할 수 있는 항목에 대한 기초통계 자료를 제시했다. 표본으로 가지고 있는 기사 1개월 치 10만 5천 건에 대해 분석이 이루어졌다. [표 3-6]은 다양한 계량 요인 중 기사 길이에 관한 기초통계 사례이다.

[표 3-6] 분류별 기사 길이에 관한 기초통계

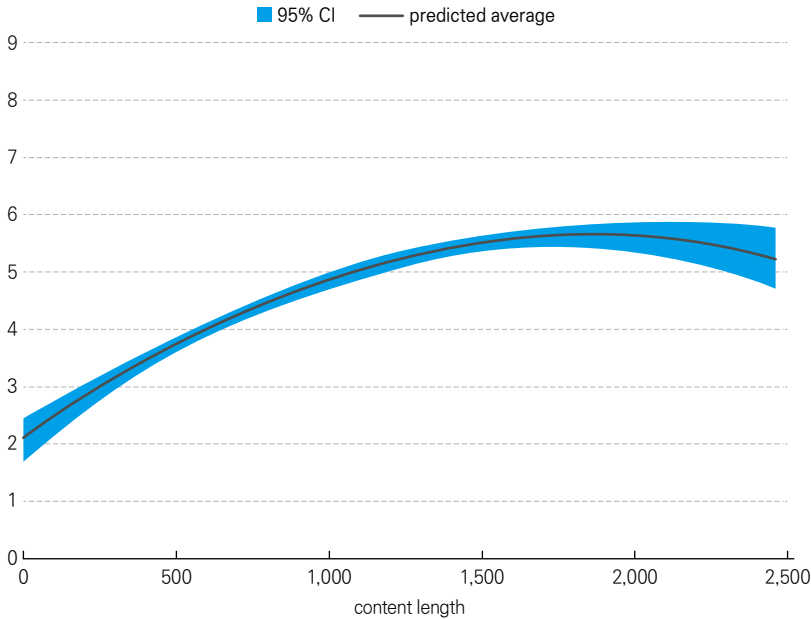
| 기사 분류 | 기사 수 | 평균 | 최대 | 최소 | 표준편차 |
|--------|--------|-------|----|--------|------|
| 정치 | 18,440 | 887 | 0 | 13,952 | 649 |
| 연예 | 20,375 | 555 | 0 | 8,882 | 418 |
| 스포츠 | 8,670 | 799 | 0 | 12,788 | 603 |
| 사회 | 76,922 | 695 | 0 | 15,364 | 589 |
| 칼럼 | 6,995 | 1,388 | 0 | 13,918 | 705 |
| 문화예술 | 12,420 | 1,086 | 0 | 12,066 | 893 |
| 라이프스타일 | 17,054 | 608 | 0 | 8,576 | 710 |
| 국제 | 9,507 | 1,005 | 0 | 8,735 | 615 |
| 교육 | 1,895 | 923 | 0 | 8,525 | 799 |
| 경제 | 46,369 | 935 | 0 | 15,672 | 686 |
| IT 과학 | 5,363 | 1,009 | 0 | 10,691 | 690 |

단순 기초통계 자료만으로는 어떤 방식으로 요인을 정해야 하는지 판단하는 것은 어려웠다. 위원회에 참가한 위원들도 전문가이지만 주관적인 판단에 기반한 의견이기 때문이었다. 따라서 기준을 정하고 점수를 부여하는 방식을 결정하기 위해 다양한 기사 분석 결과를 참고하는 과정을 거쳤다. 대학원생이 평가한 기사 1,000건 중 정치기사 105건을 중심으로 분석한 결과를 살펴보았다. 가장 논쟁적이었던 부분은 항목별 점수 산정 기준을 상대적 비율로 할 것인지, 고정적인 값으로 할 것인지에 대한 부분이었다. 하지만 상대적 비율에 따른 점수 산정 기준을 마련하는 것에 대한 반론이 있었다.

상대적 방식에 대해 문제를 제기하는 이유가 있다. 그럴 거였으면 기사 놓고 평가한 다음에 비율로 자르면 된다. 통계적으로 유의한 것 추출하는 방식은 좋은 기사 1,000개가 들어와도 나머지는 낙제 기사가 된다는 문제가 있다. (소위원회 9차, 위원 1)

이러한 논쟁은 프로젝트가 지향하는 목적과 관련되어 있었다. 저널리즘 가치를 판별하는 기준을 정하고자 하는 시도에 어려움이 있어도 절대적인 기준을 마련하고자 했던 이유는 자의적일 수 있음에도 불구하고 어떤 것이 품질 높은 기사인지에 관한 사회적 합의를 마련하려는 시도이기도 했기 때문이다. 하지만 해당 시점에 나온 기사를 상대적으로 평가하고자 한다면 시기에 따른 평가 기준이 달라질 수 있다. 물론 이에 대한 반론도 있었다. <그림3-2>는 대학원생이 평가한 기사 길이와 품질의 관계를 나타낸 결과이다. 해당 결과는 1,500자에서 2,000자까지는 품질이 높아진다고 보지만, 2,000자를 넘어가는 구간부터는 인지하는 기사 품질이 낮아지는 모습을 보였다.

<그림 3-2> 대학원생이 평가한 기사 품질과 기사 길이의 관계



이러한 분석 결과를 바탕으로 위원들이 특정한 점수 산정 기준을 만드는 것에 대한 반론은 지나치게 자의적이고 아무런 근거가 없는 방식이라는 점이었다.

한 위원은 기준을 정하고 비교할 만한 근거를 마련하는 것이 필요하다는 의견을 밝혔다. 반면, 기존 알고리즘이 활용하는 이용자 평가 기반이 아닌 저널리즘 가치를 반영하는 대안적 알고리즘을 제시하는 것이 프로젝트가 가지고 있는 목적이기 때문에 본래 지향하고자 했던 목표에 충실해야 한다는 반론도 이어졌다.

위원회에서 분포를 보고 이상적인 값을 정하는 방법. 이렇게 하면 합의에 도달하는데 근거가 없다. 근거로 삼을 게 무엇인가? 왜 기사가 1,000자 이상이면 심층성이 높은가? 저희가 논리적으로 방어할 수가 없다. 글자 수가 많아도 심층성이 계속 높아지는 것 아니다. 비교할 만한 근거가 필요하다. 근거로 삼을게 이용자 조사이다. (소위원회 9차, 위원 M)

상대적 평가는 대안적인 방식이 아니라 기존과 똑같은 것이다. 다른 플랫폼에서 만드는 것이 상대적 평가이다. 주어진 데이터 내에서 많이 본 것을 좋게 평가한다. 상대적 비율로 자르는 것은 같은 방식이다. (소위원회 9차, 위원 I)

다양한 집단의 기사 평가결과를 확인한 이후 논의를 하는 것이 바람직하다는 방향으로 의견의 합의가 이루어졌다. 고정 방식으로 품질 높은 기사를 판별하는 절대적 기준을 정하기 위해서는 여러 가지 결과를 고려하여 적절한 값을 찾기 위한 조정과정이 필요하다고 보았다.

고정 방식이라는 것도 고정 방식 기준을 정하기 위해 상대적 분석을 해보고 기준 근거를 끌어와야 한다. 자연어 처리나 인공지능 최종 목적은 기계가 처리한 결과가 사람이 하는 것과 같아지는 것이다. 여기서 대학생 모집단과 전문가가 보고 느끼는 것 두 가지가 있을 수 있다. 물론 차이는 있을 수 있다. 양자의 비율을 어떻게 조절해서 반영할지... 결과를 가지고 분석하되 지금 하신 것처럼 대학생 평가단이 일반적 피드백을 하고 전문가가 이렇게 가야 한다. 두 가지를 놓고 비율 조정하는 게 좋지 않을까. 사람들이 좋아하는 기사라고 해서 바람직하다고 할 수 없다. (소위원회 9차, 위원 N)

이러한 논의 결과에 따라 위원들이 직접 표본 기사를 평가하고 결과를 토론하며 기준을 마련하는 과정을 거쳤다. 이러한 것은 온라인에서 올바른 기사가 무엇인지 지향점을 가지는 것이 필요하다는 논의로 이어졌다. 따라서 위원회 위원들이 기사를 직접 평가한 후 다시 논의하기로 결정하였다.

위원들이 평가해 보고 집단별 차이를 봤으면 좋겠다. 지금 결과 자체를 가지고 우리 평가 기준 자체가 잘 되었다 논의하는 것은 할 수 없다. … 언론인, 학자, 정치학자, 대학원생, 일반인 집단별로 차이가 날 것 같다. 그때 가서 이야기해야 할 게 저널리즘 뉴스 역할 같은 규범적 논의이다. 어느 수준까지 할 것인가. 그 수준을 정하면 합쳐서 수용 가능한 기준을 만드는 작업이 가능할 것이다. (소위원회 9차, 위원 D)

위원회를 만들고 구성한 이유는 위원회가 지향점을 제시하는 것이다. 정확하게 입증할 수 없지만, 위원회에서 이런 기사를 디지털 공간 내에서 이러한 식으로 기사를 쓰는 것이 올바르다고 생각했기 때문에 이러한 기준에 맞춰 알고리즘을 구성했다고 밝히는 게 옳다고 생각한다. (소위원회 10차, 위원 I)

(2) 기사 평가에 기반한 상호 합의 과정

공동으로 논의할 수 있는 평가 기준을 만들어내기 위해 위원들이 정치분류 기사를 시작으로 직접 기사를 평가하는 과정을 거쳤다. 실제 기사 평가 이후 지금까지 논의가 이루어졌던 점 이외에 다양한 의견이 제시되었다. 첫 기사 평가 이후 가장 많이 언급된 주제는 다양한 특징에 따라 다른 평가 기준이 적용될 필요가 있다는 점이었다. 중앙지와 지방지 간 차이라든지 매체 속성이나 기사 포맷에 따른 차이를 고려할 필요성에 대한 부분이 언급되었다. 또한, 알고리즘으로 평가하기 위해 기계적으로 판단할 수 있는 영역이 제한적이라는 의견도 이어졌다. 그렇지만 위원들이 실제 기사를 평가하는 과정에서 그간 논의했던 부분과는 다르게 새로운 평가 기준에 관한 생각을 얻을 수 있었다.

인상적으로 기사를 평가하다가 직접 기사를 보며 평가하니 많은 것을 발견했다. 평범한 스트레이트 평가는 의미 없다. 평가가 어려울뿐더러 의미가 없다. 정치기사

속성상 형식적으로 그럴듯한데 문제 있는 기사가 많다. 예를 들면, 000 기사에서 지역지는 000 기사를 굉장히 길게 쓴다. 자세히 보면 편향적이지만 기계가 보면 좋은 기사로 볼 수 있다. 필체에 관한 기사는 형식적으로 보면 독창적인 것 같은데 검색해보니 베낀 기사이다. 기계가 평가하기 전 필터링이 안 되면 문제가 있을 수 있다. 명백하게 나쁜 것은 눈에 들어온다. 정치기사 속성상 균형성이나 다양성을 자세히 기계가 평가할 수 있을까. 계량 요인 숫자를 가지고 할 수 있는가. 기사 요건을 갖춘 것 같은데 지방지 기자가 중앙정책을 다룬 기사가 있다. 아무리 봐도 연합 기사를 베낀 것이다. 중앙지와 같이 평가해야 하나. 지방지가 중앙 국회 정국 분석이나 남북관계 분석을 쓴 부분은 신뢰도가 떨어진다. 그런 부분들 평가 레벨이 달라야 한다. 베껴 쓴 기사가 앞에 오는 경우도 있다. 그 현상이 여기서도 반복될 가능성이 있겠다. (소위원회 10차, 위원 B)

기사를 쓸 때 우리가 어떤 매체인지에 따라 포맷이 다르다. 대전지역 발행 매체는 대전 정치인들을 세세하게 쓴다. 매체 속성으로 독자가 누구인지에 따라 생산하는 것 자체가 달라진다. 그런 것들을 고려하는 게 필요하다. 그걸 할 수 있을까? 기사 포맷이 무엇인가. 인터뷰 형 기사인 경우, 피쳐 기사인 경우, 스트레이트인 경우 같은 대상으로 접근하지만 기사를 풀어내는 서술법은 완전히 달라진다. 인터뷰 기사인데 다양성 균형성으로 접근할 수 있는가? 000 후보 이야기에 다른 건 들어가지도 않는다. 기사 포맷이 달라지는 경우 동물로 비교할 수 있을 것인가. 어떤 지표들을 보니 기사 표본이 너무 작다. 어떤 것은 해당하는 기사가 몇 건 안 되는 대단히 이례적인 기사다. 어떤 속성이나 지표들은 측정하기 쉽고 어떤 것은 어렵겠다는 사실을 깨닫게 하는 유의미한 작업이라고 생각한다. (소위원회 10차, 위원 A)

결국, 문제는 평가 기준을 위해 논의했던 내용과 다르게 발견된 기사 사례가 정말 예외적인지, 일반적인 관행으로 자리 잡은 부정적인 기사인지를 판단하는 것이 필요하다는 점이었다. 예를 들면, 출처를 정확히 인용하는 것은 긍정적인 것으로 간주했었으나, 실제 기사 평가에서는 따옴표 인용만을 통해 스트레이트와 큰 차이를 보이지 않는 단순 기사를 생산하는 사례가 있었다.

긍정적일 것으로 논의했던 측정방식이 실제 기사를 확인해보니 부정적으로 나타날 수 있을 것 같다. 예를 들면, 따옴표는 출처 인용을 측정하는 방식이지만 따옴표와 접속사만을 사용하여 스트레이트와 유사한 형태의 단순 기사를 생산하는 사례가 있다. (소위원회 10차, 위원 O)

이후 반복적으로 이루어진 평가에서도 비슷하게 개념 적용에서 차별성이 필요하다는 내용과 각 개념을 측정하는 계량 요인을 어떤 기준에서 볼 것인지에 관한 토론이 이어졌다. 이러한 과정은 기사 품질을 평가하는 기준을 마련하기 위한 것이기도 했지만, 현재 온라인 기사 생산 관행에 대해 다시 한번 살펴보고 프로젝트가 지향해야 할 바를 다시 확인하는 과정이었다.

기사를 읽으며 평가해 보니 현장에서 독특한 기사를 쓰지 않고 있구나 하는 반성을 했다. 특종 말고는 별로 고민을 안 한다. 사실 비슷한 소재를 가지고 쓰다 보니 하나 마나 한 기사다. 다양성이 부족한 것 반성한다. 나쁜 기사는 확실히 나쁘다. 표절이나 간접 취재와 같은 부분을 걸러내야 한다고 생각했다. 관행에 의한 평가가 알고리즘으로 평가하기 쉽지 않겠고 차별화도 쉽지 않겠다는 생각을 했다. 알고리즘을 잘 만든다고 하더라도 기사가 비슷비슷해서 차이가 없지 않을까 생각했다. (소위원회 11차, 위원 H)

스포츠 연예 기사를 평가해 보니 가지고 있는 10가지 기준이 분야에 따라 차별 적용되어야 하지 않을까 싶었다. 정치기사에서 정치적 균형성이나 의견의 다양성이 균형을 이루고 있는 것과 전문적인 한쪽의 주장을 심층적으로 들어가는 것을 같이 놓고 볼 수 없다. 많은 전문가가 들어와야 꼭 전문성 있는 기사인가. 그런 부분에 대해서 가중치를 어떻게 줄 것인지 고민할 필요가 있음. 평가하면서 나쁜 것을 먼저 제외하는 것으로 힘을 모아야 하지 않을까. (소위원회 11차, 위원 P)

평가결과 중에서 불일치가 큰 기사를 중심으로 재논의가 이루어졌다. 이러한 과정을 거치며 합의한 주요 내용은 기사 분류별·유형별 차별적 기준 적용, 부정적 기사를 걸러낼 수 있는 평가접근, 측정 변인 적용 방식에 관한 것이었다.

전반적으로 기사 분류별 차이가 있다는 사실을 인지한 가운데 익명 취재원에 관한 부정적 평가, 감정적 수식어 제한, 간접 취재에 대한 부정적 평가, 출처에 대한 긍정적 평가가 합의된 가운데 특정 기사 분류에 따라 허용할 여지가 있는 부분을 합의했다. 다양한 기준을 고려하는 것이 어려워 부정적인 요소에 집중하여 전반적인 합의사항을 기반으로 판별하는 방향으로 합의가 이루어졌다. 모든 위원이 오랜 기간 토론하며 얻은 결과는 긍정적인 부분은 다양하고 합의가 어렵지만, 저널리즘에서 부정적인 요소가 무엇인지는 상대적으로 쉽게 합의할 수 있다는 점이었다. 기사 유형, 언론사 유형, 보도자료 여부에 따른 차이도 논의되었으나, 이러한 결과를 적용하기 위한 기술적인 개발과정이 더 요구되는 부분이었다.

[표 3-7] 평가에 관한 합의사항

| 기사 분류 | | 합의사항 | 논의사항 |
|-------|--------|--|---|
| 분류 | 정치 | <ul style="list-style-type: none"> • 익명 취재원 • 감정적 수식어 비사용 | <ul style="list-style-type: none"> • 취재원 수와 다양성 • 기사 길이와 인용 비율 |
| | 경제 | <ul style="list-style-type: none"> • 보도자료로 인해 차별성 낮음 | <ul style="list-style-type: none"> • 익명 취재원 허용 여부 • 균형성과 다양성 판단 |
| | 사회 | <ul style="list-style-type: none"> • 간접 취재에 대한 부정평가 | - |
| | 국제 | <ul style="list-style-type: none"> • 외신 기사 인용 출처 • 기사 길이에 대한 긍정 평가 | - |
| | 스포츠 | <ul style="list-style-type: none"> • 스포츠 분야는 간접 취재 | - |
| | 연예 | <ul style="list-style-type: none"> • 평가할 요인이 없음 | <ul style="list-style-type: none"> • 소셜미디어 인용 평가 |
| 유형 | 기사 유형 | <ul style="list-style-type: none"> • 스트레이트, 피쳐, 인터뷰 차이 | <ul style="list-style-type: none"> • 기술적으로 사실 칼럼 분류 불가 • 단일 출처 전문적 취재원 평가 |
| | 언론사 유형 | <ul style="list-style-type: none"> • 지방지와 중앙지 평가 | - |
| | 보도자료 | <ul style="list-style-type: none"> • 보도자료 클러스터링 후 평가 | - |
| 접근방식 | 부정적 접근 | <ul style="list-style-type: none"> • 부정적 기사는 합의되는 경향 • 긍정적 기사 평가 합의는 어려움 | <ul style="list-style-type: none"> • 부정적인 요인 점수 부여 기준 |

(3) 다양성의 적용과 반복적 조정작업

합의한 부분을 바탕으로 기사에 대한 평가 기준 초안이 마련되었다. 예를 들면, 기자 명의 경우 실명 기자 명과 이메일 주소가 모두 있는 경우에는 1점, 실명 기자 명만 있는 경우 0.8점, 비실명 기자 명 또는 이메일 주소만 있는 경우에는 0점, 아무것도 없는 경우에는 -1점과 같이 점수가 부여되었다. 이러한 평가 기준의 적용에서 기자 명은 기사 분류나 매체 유형에 상관없이 모두 같게 적용하였다. 반면 저널리즘 가치에서 기자 명은 투명성과 사실성에만 연결되어 있다고 보고 해당 가치를 반영하는 계량 요인으로 판단하였다.

기사의 길이, 인용문 수, 제목의 길이, 제목의 문장부호, 수치 인용 수, 이미지 수, 평균 문장의 길이, 제목에 사용된 부사 수, 문장당 평균 부사 수, 기사 본문의 인용문 비율 모두 각 기준에 따른 평가 안이 마련되었다. 평가 기준의 결정은 위원들이 반복적인 기사 평가 과정을 통해 경험한 내용과 표본 기사 10만 5천 개에 대한 통계치를 기준으로 결정하였다. 이러한 평가 기준을 통해 최종적으로 산출되는 기사에 대한 평가 점수는 저널리즘 가치에 따라 다를 수 있으므로 평가 점수와 개념별 가중치의 함수로 결정하는 방식을 적용했다. 이는 저널리즘 가치와 측정 요인을 연결하는 과정에서 각 측정 요인이 어떤 가치를 더 잘 반영하고 덜 반영하는지에 대한 논의과정에서 합의된 것이었다.

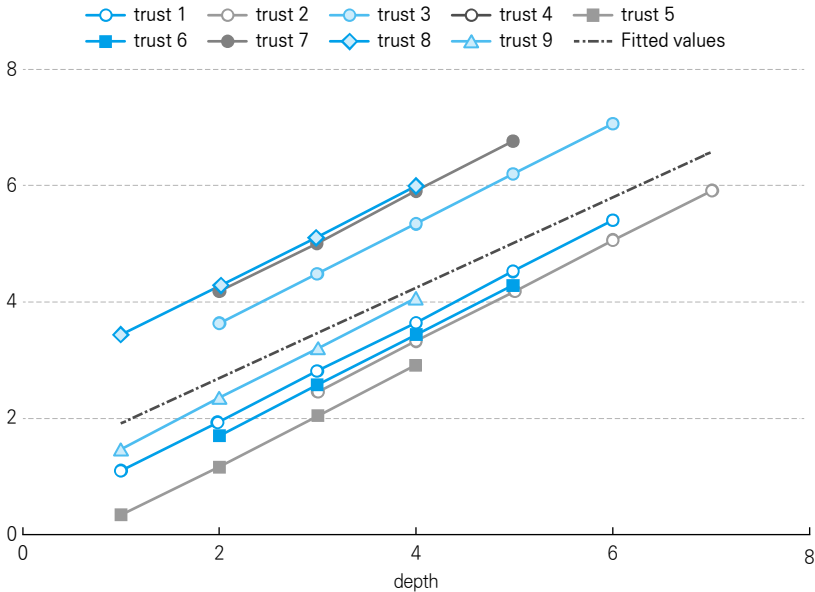
사회, 스포츠, 연예를 봤는데 기준이 장르별로 달라야 한다는 것. 사회 기사는 이런 가중치가 들어가야 하고, 스포츠는 이런 게 들어가야 한다. (소위원회 11차, 위원 H)

기사 평가를 계속 반복하면 정리될 것 같다. 처음 우리가 했던 가치 개념 정리를 다시 해야 한다. (소위원회 11차, 위원 D)

실제 위원들이 평가한 기사 점수를 바탕으로 살펴본 결과도 각 위원에 따른 차이를 확인할 수 있었다. 정치분류 기사를 평가한 결과를 살펴보면 각 위원이 인지하는 기사의 전반적인 품질과 개별 가치가 품질에 이바지하는 정도에서 개인

간 차이가 드러난다는 사실을 발견할 수 있었다. <그림3-3>에서 확인할 수 있는 것처럼 기사 품질과 심층성은 정적 상관관계를 보였다. 즉, 심층성이 높아질수록 기사 품질이 증가하는 형태를 보였다. 하지만 심층성이 전체 품질에 어느 정도 이바지하는지는 개별 위원마다 인식의 차이를 보였다. 위원 5 같은 경우는 전반적으로 심층성이 기사 품질에 이바지하는 부분을 낮게 보았지만, 위원 7은 심층성이 기사 품질에 이바지하는 부분이 높았다. 위원회에서 결정한 11가지 저널리즘 가치가 모두 중요하지만 어떤 기사에서는 특정 가치가 더 중요하고, 다른 기사에서는 그렇지 않을 수 있기 때문이었다.

<그림 3-3> 기사 품질과 심층성에 대한 개별 위원 평가결과



우리가 가지고 있는 11개의 요인. 이 가운데서도 주제별로 더 중요한 것은 가중치임. 11개 요인이 다 중요한데 이 기사에서는 이 요인이 중요하다는 것을 나타내기 위해 기사 분류별 평가를 한 것이다. (소위원회 10차, 위원 H)

저널리즘 가치와 계량 요인의 관계는 양쪽을 연결하기 위한 과정에서 위원들이 평가한 점수에서 실제 기사 평가 후 적절한 가중치라고 생각하는 범위를 정한 초안을 만들었다. 이러한 방식을 적용한 것은 위원회가 판단한 높은 품질의 기사 기준을 바탕으로 설명 가능한 저널리즘 가치 기반 알고리즘을 제시하는 것이 목표였기 때문이다. 가중치에 대한 정확한 값을 위원들이 결정하는 것이 아니라 개념과 관계가 있는지 없는지, 방향성은 어떤지에 대한 위원 의견을 종합하여 [표 3-8]에서 보는 것과 같은 가중치 초안이 제시되었다.

[표 3-8] 저널리즘 가치와 계량 요인 관계 기반 가중치 안

| | 균형성 | 다양성 | 독이성 | 독창성 | 사실성 | 심층성 | 유용성 | 중요성 | 투명성 | 선정성 |
|-------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 기사 명 | 2~3 | 0~1 | 0~1 | 3.5~4.5 | 3.5~4.5 | 3.5~4.5 | 2.5~3.5 | 1.5~2.5 | 3.5~4.5 | 3.5~4.5 |
| 기사 길이 | 3~4 | 4~5 | 0~1 | 3.5~4.5 | 3.5~4.5 | 4~5 | 2.5~3.5 | 3.5~4.5 | 3~4 | 0 |
| 인용 수 | 2~3 | 2.5~3.5 | 0~1 | 3.5~4.5 | 3.5~4.5 | 3.5~4.5 | 2~3 | 3.5~4.5 | 3.5~4.5 | 0 |
| 제목 길이 | 0 | 0.5~1.5 | 1~2 | 0 | 0 | 0 | 0 | 0 | 0 | 3.5~4.5 |
| 제목 문장부호 | 1.5~2.5 | 0~1 | 1~2 | 3~4 | 0~1 | 0~1 | 0 | 0 | 3.5~4.5 | 3.5~4.5 |
| 수치 인용 수 | 0 | 1~2 | 0.5~1.5 | 1~2 | 0.5~1.5 | 0.5~1.5 | 1~2 | 1~2 | 0.5~1.5 | 0 |
| 이미지 수 | 0 | 1~2 | 1.5~2.5 | 1.5~2.5 | 1.5~2.5 | 1~2 | 1~2 | 0.5~1.5 | 1~2 | 0 |
| 평균 문장 길이 | 0 | 0.5~1.5 | 1.5~2.5 | 0 | 0 | 0~1 | 0~1 | 0 | 0 | 0 |
| 제목 부사 수 | 0.5~1.5 | 0~1 | 1.5~2.5 | 0 | 0.5~1.5 | 0 | 0 | 0 | 0 | 3.5~4.5 |
| 문장당 평균 부사 수 | 1~2 | 0~1 | 0.5~1.5 | 0 | 1.5~2.5 | 0~1 | 0 | 0~1 | 0~1 | 3.5~4.5 |
| 인용문 비중 | 1~2 | 0.5~1.5 | 0~1 | 0~1 | 1~2 | 1~2 | 0~1 | 0 | 0~1 | 3.5~4.5 |

특정한 가중치의 경우 인위적으로 가중치 범위를 제한할 필요성이 있었다. 아무런 제약 없이 위원회에서 한 기사 평가결과 데이터에 가장 적합한 값으로 맞추는 방식은 기존 알고리즘에서 활용하는 이용자 기반 데이터를 통한 추정값에 의존하는 것과 별다른 차이가 없었기 때문이다. 가치 기반의 설명할 수 있는 알고리즘을 제시하는 것이 목표였기 때문에 인위적으로 가중치의 범위를 제한했고, 이를 바탕으로 기계 학습을 통해 가중치를 조정하는 과정을 거쳤다. 최종적으로 결정된 가중치는 [표 3-9]와 같았다.

[표 3-9] 저널리즘 가치와 계량 요인 관계 기반 가중치 결과

| | 균형성 | 다양성 | 독이성 | 독창성 | 사실성 | 심층성 | 유용성 | 중요성 | 투명성 | 선정성 |
|-------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 기자 명 | 2.996 | 0.998 | 0.001 | 4.494 | 4.493 | 4.496 | 3.494 | 2.495 | 4.498 | 4.491 |
| 기사 길이 | 3.002 | 4.994 | 0.003 | 4.492 | 3.503 | 4.995 | 3.498 | 3.503 | 3.003 | 0 |
| 인용 수 | 3.000 | 2.501 | 0.001 | 3.501 | 3.501 | 3.501 | 2.001 | 3.500 | 4.500 | 0 |
| 제목 길이 | 0 | 0.500 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 3.5 |
| 제목 문장부호 | 1.501 | 0 | 1.002 | 3.09 | 0.001 | 0 | 0 | 0 | 3.619 | 3.501 |
| 수치 인용 수 | 0 | 1.953 | 1.354 | 1.823 | 0.502 | 1.336 | 1.956 | 1.002 | 1.454 | 0 |
| 이미지 수 | 0 | 1 | 1.500 | 1.501 | 1.500 | 1 | 1 | 0.500 | 1 | 0 |
| 평균 문장 길이 | 0 | 0.500 | 1.500 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 제목 부사 수 | 0.501 | 0 | 2.466 | 0 | 0.500 | 0 | 0 | 0 | 0 | 3.500 |
| 문장당 평균 부사 수 | 1.000 | 0 | 0.500 | 0 | 1.500 | 0 | 0 | 0 | 0 | 3.500 |
| 인용문 비중 | 1 | 0.500 | 0 | 0 | 1 | 1 | 0 | 0 | 0.001 | 3.500 |

위원회에서 논의한 결과는 우리가 가지고 있는 기사 평가 기준을 단 하나의 기준으로 환원하는 것이 불가능하다는 사실을 보여주었다. 따라서 알고리즘을 공개했을 때, 이용자가 스스로 원하는 저널리즘 가치와 가중치를 선택하여 적절한 기사 분류 결과를 확인할 수 있도록 만드는 인터페이스를 적용하여 알고리즘 결과를 공개했다. 적용한 결과는 이용자 관점에서 알고리즘 다양성을 제시했지만, 알고리즘 정확도를 높이기 위해 반복적인 조정작업이 필요했다. 결국, 알고리즘은 사람들이 보기에 그럴듯하다고 판단하는 결과를 따라 하는 것이었기 때문이다.

3. 뉴스 트러스트 알고리즘 개발 결과

1) 배열 알고리즘 적용을 위한 시스템 개발

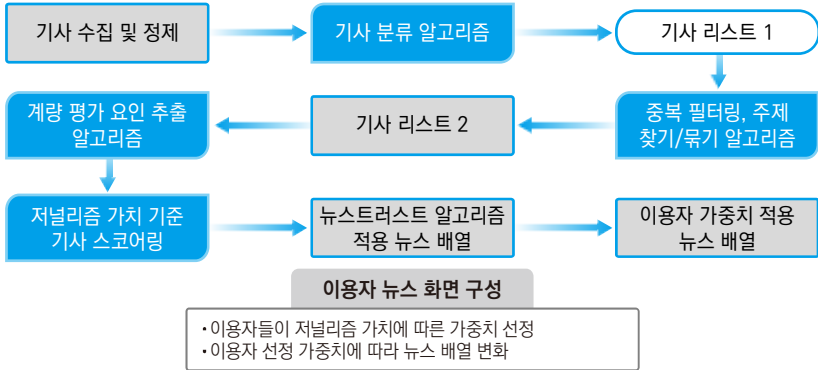
(1) 전체 개발 모델링

뉴스 트러스트 위원회의 논의와 함께 이를 기술적으로 구현하기 위한 절차도 바로 시작됐다. 먼저, 전체 시스템에 대한 설계가 우선이었다. 모델링 과정에서 전체 기사가 아닌 클러스터링된 기사들을 대상으로만 배열을하기로 결정했다. 논의 과정에서 기사 배열에 있어서 완전한 정답은 없다는 데 의견이 모아져 비슷한 주제의 기사들 중 점수에 따라 배열하는 것으로 합의됐기 때문이다. 또한, 뉴스 트러스트의 결과물은 목표를 향해 가는 과정이라는 점을 감안해 이용자들이 저널리즘 가치별로 가중치를 설정할 수 있는 기능을 추가하기로 결정했다. 이를 흐름도로 표현하면 <그림 3-4>와 같다.

먼저, 배열 알고리즘의 적용하고 테스트할 대상이 되는 기사들을 수집하고 이를 정제하는 작업을 실시했다. 2차년도까지는 빅카인즈 수집 DB 데이터 내 기사들을 수집 및 사용해서 분석 작업을 진행했고, 3차년도에는 최신 기사 적용을 위해 포털 기사 크롤링 데이터를 활용해서 알고리즘 테스트를 진행했다. 기사 수집 및 정제가 완료된 이후에는 해당 기사들을 뉴스 트러스트 위원회가 정한 11가지 분류로 자동 분류하는 과정을 거쳤다. 분류 이후 정리된 기사들을 대상으로 중복 기사들을 제거하고 비슷한 주제의 기사끼리 묶는 클러스터링 작업을 진행하였다. 클러스터링은 1일을 기준으로 실시하여 수집한 기사들이 하루 단위로 비슷한 기사들끼리 묶이도록 하였다. 클러스터링된 기사들을 대상으로 기사명 추출, 개체명 인식, 인용문 추출 등 뉴스 트러스트 위원회가 결정한 계량요인의 추출 작업을 진행하였다. 추출된 요인을 뉴스 트러스트 위원회가 정한 규칙에 따라 점수를 매겼고 그 점수에 따라 비슷한 주제로 묶인 기사들 간에 배열 위치를 정했다. 이용자들은 하루 단위로 묶인 클러스터 내 기사들이 부여받

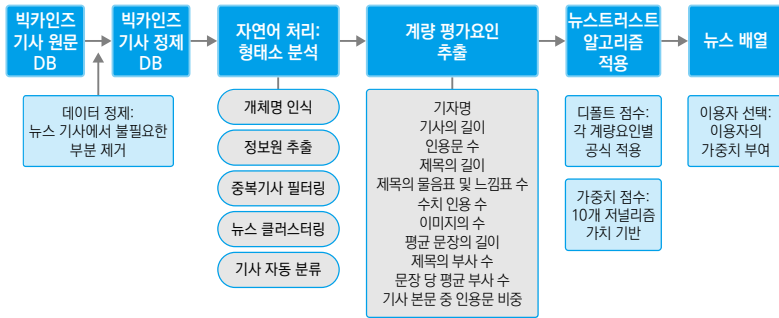
은 점수에 따라 배열된 화면을 볼 수 있으며, 최종적으로 배열에 적용된 저널리즘 가치별로 가중치를 달리 하여 배열 위치를 달리하여 볼 수 있도록 모델링하였다.

〈그림 3-4〉 뉴스 트러스트 전체 개발 흐름도 구성

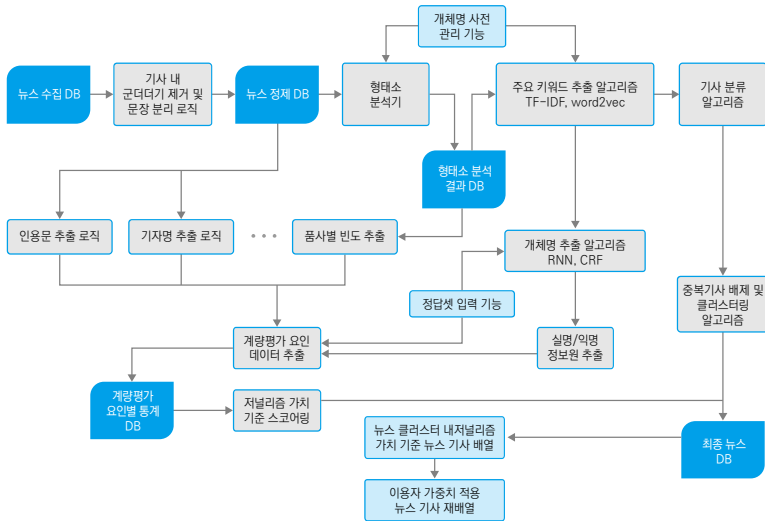


모델링에 따른 흐름도는 〈그림 3-4〉와 같지만, 실제 개발 과정에서는 흐름도 순서와 별도로 동시에 진행되어야 하는 과정들이 있었다. 예를 들어, 기사의 자동 분류를 위해서는 계량 요인 추출 등에 적용되는 자연어 처리 적용 등이 동시에 이루어져야 했다. 이에 따라 실제 개발은 〈그림 3-5〉와 같은 과정으로 이루어졌다. 빅카인즈 기사 DB에서 원문을 수집한 후 데이터 정제를 통해 정제 DB를 별도로 구축한 후 이 DB 내 기사들을 대상으로 형태소 분석 등 자연어 처리 과정을 거쳤다. 개체명 인식, 정보원 추출, 자동 분류, 클러스터링 등의 작업을 동시에 진행했다. 자연어 처리 결과를 바탕으로 기사 명, 기사의 길이, 인용문 수, 제목의 길이, 제목의 물음표 및 느낌표 수, 수치 인용 수, 이미지의 수, 평균 문장의 길이, 제목의 부사 수, 문장당 평균 부사 수, 기사 본문 중 인용문 비중 등 뉴스 트러스트 위원회에서 결정한 계량 요인 11가지를 각 기사별로 추출하였다. 이후 위원회의 공식에 따라 각 기사별로 점수를 부여하여 최종적으로 이용자 화면에서 배열하였다. 이렇게 하여 최종적으로 구성된 시스템은 〈그림 3-6〉과 같다.

〈그림 3-5〉 뉴스 트러스트 전체 개발 과정



〈그림 3-6〉 뉴스 트러스트 전체 시스템 구성도



(2) 기사 수집·정제 및 기사 명 추출

전체 개발 모델링이 이루어진 후 가장 먼저 한 작업은 분석 및 적용 대상 기사를 수집하고 수집한 기사를 분석이 가능한 형태로 정제하는 작업이었다. 2015년 1월 1일부터 2016년 6월 30일까지 빅카인즈 수집 DB에 있는 뉴스 기사 2,818,676건을 수집하였다. 분석에 활용하기 위해 기사 내 광고 및 부가정보를

제외하고 본문만 추출하는 작업을 진행했다. 상단 제목 이후부터 하단 기자 명 이전까지만 본문으로 정의하고 그 외 부분은 제거하였다. 예를 들어, 아래와 같이 기사 내에 관련 기사 및 부가 정보가 있을 경우 이를 모두 제거하였다.

- ☞ [산으로 가는 구조조정 · 끝] 뼈를 깎고 살 도려내서 정상화시켜라
- ☞ 수출 17개월째 '후진' 탈경대는 '성장 엔진'
- ☞ 유한킴벌리, 생리대 값 인상 논란 속 '신제품' 가격 ↑
- ☞ "페이스북 같은 대기업 한국선 나올 수 없어" [OO닷컴 바로가기 | 소셜 서유기 | 모바일 웹]

빅인즈에 지정된 전송 형식이 있었지만 각 언론사마다 각기 다른 기사 쓰기 유형이 있어서 언론사별로 특성을 파악하여 본문 내용만 정제하는 작업을 진행했다. 기사 마지막에 포함된 공백들 즉, 눈에 보이지 않는 띄어쓰기 기호나 유니코드 공백문자 등을 자동으로 제거하고, 기자 명 이전 부분에서도 불필요한 문자들 제거하였다. 이와 함께, 문장 띄어쓰기 형식도 통일하기 위해 문장 마침표(형태소 분석상 SF)로 끝나면 하나의 문장으로 보며, 줄바꿈이 두 번 이상 이어지면 하나의 문장이 끝난 것으로 봐 처리하였다. 문장이 문장 마침표로 끝난 뒤에, 빈 칸 없이 새로운 문장이 시작되는 경우와 (?), (!), () 등 괄호로 묶인 문장 마침표는 무시하였다. 문장 속에 "(Double Quote)"나 "(Single Quote)"로 묶인 경우, 문장으로 간주하지 않았다. 분리된 문장은 앞 뒤 공백을 제거하며, Quote가 닫히지 않은 경우, Quote 단위로 문장을 분리하지 않았다.

이러한 작업을 통해 기사 내 본문 외 부수적인 광고나 연관 기사의 경우 대부분 자동으로 제거할 수 있었으나 본문 내 소제목, 중간제목 등의 경우 일관된 형태가 없어 제대로 처리하지 못했다. 언론사별로 일정한 쓰기 형식이 있으면 찾아서 처리하였으나 일정 형식에 어긋날 경우에는 미처리됐다. 또한, 기사 본문 중간에 본문과 상관없는 내용이 등장할 경우에는 처리가 불가능하였다. 이로 인해 실제 본문 기사 길이를 측정함에 있어서 몇몇 기사의 경우 약간의 오류가 생길 수 있다.

기사 데이터 정제 후에는 본문 내 표기돼 있는 기자 명을 자동으로 추출하는

작업을 진행했다. 빅카인즈 DB에는 기자 명을 별도 필드로 입력하게 되어 있으나, 별도 필드에 기자 명이 전혀 입력되어 있지 않은 언론사가 10여 곳 정도 되었고, 표기가 되어 있더라도 제대로 입력되지 않은 경우도 많았다. 따라서 기사 본문 내에 기자 명이 입력되어 있는 부분을 추출해 기자 명을 별도로 등록하였다. 기사 마지막 줄에 기자 명으로 보이는 형식으로 끝나는 경우, 예를 들어, ‘~기자’ 혹은 ‘~기사 <이메일 주소>’, ‘기자’가 아닌 다른 형태의 직위(예를 들어, 리포터, 논설위원, 객원기자 등)를 언론사별로 입력 유형을 파악하여 자동 추출 및 등록하였다.

데이터 정제 작업은 한 번 설계하여 적용하면 끝나는 작업이 아니라 데이터가 변경되면 다시 데이터 분석 작업 후에 적용 과정을 개선해야 하는 작업으로 끝이 없는 작업이다. 1차로 데이터 이관이나 크롤링 등 작업을 통해 원본 기사들을 수집하는 작업을 진행하고 이후로 기사 제목과 본문을 분리하는 작업, 기사 관련 메타 정보들을 추출하는 작업들을 수행해야 한다. 빅카인즈 수집 DB 데이터의 경우 각 언론사에서 뉴스 기사를 입력하면서 기자 명 같은 중요한 메타 정보가 누락되어 입력되는 경우도 많이 확인되었다. 데이터 수집과 정제하는 작업은 전체 작업 중에서 시간과 노력을 많이 들여야 하는 작업이지만 실제 성과로 측정하기 어려운 영역이기도 하고, 한번 잘 정리해서 작업을 한다고 해도 다음에 적용하려고 할 때 다시 처음부터 작업해야 하는 부분이기도 해서 여러모로 어려움이 많은 작업이었다.

(3) 한국어 형태소 분석(자연어 처리)

뉴스 트러스트 프로젝트는 기사 표현을 분석, 문장 개수 파악, 불필요한 문장을 걸러내는 등 다양한 방식으로 텍스트를 분석하는 작업이었기에 이를 위해서는 한글 형태소 분석이 필요했다. 한글 형태소 분석기는 자체적으로 개발하지 않고 오픈소스로 개발된 기존 형태소 분석기 꼬꼬마, 한나눔, Mecab-ko, 코모란의 성능과 품질을 비교해 선택했다. 이 중 ‘Mecab-ko’이 품질은 비슷한데 성능 면에서 다른 형태소 분석기와 비교하여 수행 속도가 다소 빠른 것으로 판단돼

‘Mecab’을 선택해서 사용했다. 2차년도에 뉴스 트러스트에 필요한 핵심 기능에 대해서 높은 품질의 결과물을 뽑아낼 수 있는 형태소 분석기를 직접 개발해보기 위해서 개발을 진행해보았으나 오픈소스 형태소 분석기 정도의 품질을 구현하는데 실패해서 3차년도까지 계속 Mecab을 사용하였다. 다만 그대로 사용한 것은 아니고 Mecab의 품질을 개선하기 위한 작업을 여러 방면에서 시도하고 적용하였다. 기본적으로 조합된 명사 표현 추가, 고유명사 분리 기능 개선, 불필요 정보 제거, 분류된 샘플 관리, 품사에 따른 TF-IDF 가중치 부여 작업 등의 개선 작업을 진행하였다.

‘Mecab-ko’는 오픈 소스 형태소 분석 엔진인 일본의 ‘MeCab’을 사용한 한국어 형태소 분석을 위한 프로젝트로, 21세기 세종계획 모든 현대 말뭉치에서 50문장씩을 추출하여 학습에 이용하였다(총 23,615 문장). 뉴스 트러스트에서는 오픈소스로 공개된 라이브러리에 뉴스 트러스트 자체적으로 기능을 추가한 라이브러리를 사용했다. 그중 한국어 형태소 분석 성능 향상을 위해 별도의 사전 관리 기능을 추가하였다. 성씨(김, 이, 박 등), 지위 및 호칭(교수, 기자 등), 사람명사(질문자, 고용주 등), 지명(도시명 등), 지명 결합 명사(시, 구, 동 등), 장소, 조직명(청와대, 국가정보원 등), 조직 명사(팀, 본부 등), 국가명 등을 보다 효율적으로 인식하기 위한 방안이었다. <그림 3-7>과 같이 사전 입력 및 관리 기능을 추가하여 정확한 개체명 인식이 가능하도록 기능을 개선하였다.

정제된 기사 본문에서 문장들을 분리하고 각 문장 별로 ‘Mecab’을 활용하여 품사를 태깅하고 태깅된 결과를 다시 문장별로 조합하였다. 형태소는 실질 형태소를 기준으로 파악하며, 관형사는 형용사로 판단하였다. 이외 구체적 내용에 대해서는 국어학 전문가의 자문을 받아 태깅하였다. 태깅된 내용의 검증을 위해 <그림 3-8>과 같은 시각화 페이지를 만들었다. 특정 단어 위에 마우스를 올려놓으면 태깅된 품사의 목록이 나오도록 해 제대로 형태소 분석이 이루어졌는지를 확인할 수 있도록 했다.

〈그림 3-7〉 뉴스 트러스트 개체명 사전

명사

어절

명사 입력

개체명 인식 메타 정보

☐ 성명 (예: 김대중)

☐ 이름 (예: 대중,영삼)

☐ 성씨 (예: 김,이)

☐ 지위 및 호칭 (예: 씨,교수,회장)

☐ 사람 명사 (예: 지도자,간담자)

☐ 지명 (예: 서울,부산)

☐ 지명 결합 명사 (예: 시,도)

☐ 장소 (예: 지역,동해안)

☐ 조직명 (예: 공정거래위원회)

☐ 조직 명사 (예: 위원회,은행,회사)

☐ 국가명 (예: 한국,미국)

추가

기타 메타 정보

메타 정보를 선택해주세요.

〈그림 3-8〉 뉴스 트러스트의 형태소 분석 시각화 화면

mecab-ko

0

[송청정보 이청규기자] 송북도 가 구재역 AI 청정 지역 을

1

송북도 는 지난해 10월 부터 운영 했 던 구재역 AI 특별방역 대책기간 이 31일 로 종료 됐다고 이날 밝혔다

2

위기관리 단계 도 진행 주 의 에서 관심 단계 로 하향 조정 됐 다

hannanum

0

[송청정보 이청규기자] 송북도 가 구재역 AI청정 지역 을 잘 가 지

1

지가 이 내 있다 송북도 는 지난해 10월 부터 운영 하 있던 구재역 AI 특별방역 대책기간 이 31일 로 종료 되 었다 고 이날 밝히 었 다 위기관리 단계 도 진행 주 의 에서 관심 단계 로 하향 조정 되 었 다 송북도 는 그러나 구재역 AI 상시 방역태세 는 계속 유지키 로 하 있다 특별방역기간 동안 전국적 으로 구재역 21건 조방인출방역자 26건 이 발생 하 있다

kkma

0

[송 청 정보 이 청 규 기 자] 송 북 도 가 구 재 역 AI 청 정 지 역 을 잘 가 지 지가 내 었 다 ETD-관청청 전상 이의

1

송 북도 는 지난해 10 월 부 터 운영 하 있 던 구 재 역 AI 특 별 방 역 대 책 기 간 이 31 일 로 종 료 되 었 다 고 이 날 밝 히 었 다

2

위 기 관 리 단 계 도 현 행 주 의 에서 관 심 단 계 로 하 향 조 정 되 었 다

twitter

0

[송청 정보 이 청 규 기 자] 송북도 가 구재역 AI 청정 지역 을 잘 가 지 지가 내 다

1

송북도 는 지난해 10 월 부 터 운영한 단 구재역 AI 특별 방역 대책 기간 이 31 일로 종료 됐 다고 이 날 밝히 다

2

위 기 관 리 단 계 도 현 행 주 의 에서 관 심 단 계 로 하 향 조 정 됐 다

또한, 사전 등으로 해결이 안 되는 다양한 구문에 대한 정확한 인식을 위해 사람이 직접 입력한 결과를 기계가 학습하도록 해 형태소 분석 결과 정확도를 높일 수 있게 했다. 예를 들어, 위원회의 결정에 따라 무주체 술어(~한 것으로

알려졌다 등) 인식을 자동으로 해야 했는데, 무주체 술어 사전이 따로 있는 것이 아니라 사람이 여러 사례를 입력하여 기계가 학습하도록 했다.

(4) 기사 자동 분류

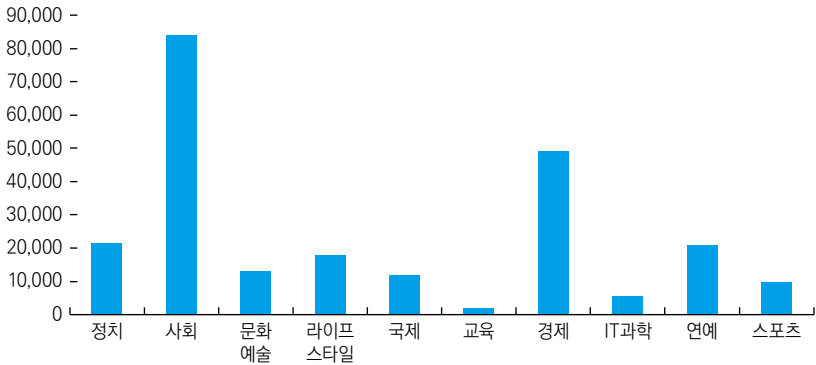
뉴스 트러스트 위원회는 기사의 분류를 IT과학, 경제, 교육, 국제, 라이프스타일, 문화예술, 사회, 스포츠, 연예, 정치, 사설·칼럼 등 총 11개로 정했다. 이러한 분류체계는 언론사들의 자체적 분류와 정확히 일치하지 않기 때문에 기사의 내용을 분석하여 뉴스 트러스트의 분류에 맞게 자동으로 재분류하였다. 1차년도에 기사 분류 작업은 기사별로 TF-IDF를 추출한 후 나이브 베이즈(Naive Bayes) 알고리즘을 적용하였다. 나이브 베이즈 알고리즘은 텍스트 분류에 주로 사용되는 것으로 문서를 여러 범주(예: 스팸, 스포츠, 정치 등) 중 하나로 판단하는 일반적인 방법 중 하나다. 단일 알고리즘을 통한 훈련이 아닌 일반적인 원칙에 근거한 여러 알고리즘을 활용하여 훈련된다. 하지만, 정답셋 데이터가 너무 부족하다는 문제가 있었다. 사람이 직접 기사를 분류해서 정답셋을 만들어 진행하다보니 정답셋을 충분히 확보하지 못했고 이를 기준으로 전체 기사에 대한 분류를 유추하는 작업은 정확성이 낮은 결과물을 가져왔다.

이를 개선하기 위해서 2차년도에는 언론사에서 1차적으로 분류한 분류정보를 뉴스 트러스트의 분류와 맞게 조정한 후 이를 정답셋으로 사용하고 분류가 없는 미분류 기사나 일반적인 분류로 분류가 되어있지 않은 기사에 대해 기계 학습을 통한 분류 유추 작업을 진행하였다. 17,000개의 기사를 대상으로 분류에 대한 기계 학습을 3차례 실시했다. 뉴스 기사 본문을 형태소 분석기를 사용해서 고유명사, 일반명사, 동사, 형용사만 사용하고 전체 기사 본문의 텍스트를 초성, 중성, 종성으로 나눠 사용하였다. 이 후 약 3,000개 데이터에 대해 정확도를 검증한 결과 75%의 정확도를 기록했다. <그림 3-9>는 기사를 자동으로 분류한 결과다.

결과적으로 정답셋 데이터가 충분히 많이 확보되어야 만족할 만한 결과를 확보할 수 있고 기존에 확인된 분류 결과물을 정답셋으로 추가하는 작업을 꾸준히 진행할 수 있다면 결과의 정확성을 더 높일 수 있을 것으로 보인다. 하지만,

분류에 대해서는 완전한 합의가 이루어지지 않고 애매한 경우가 많아서 100% 완전한 결과를 얻기는 어려울 것으로 판단된다.

〈그림 3-9〉 기사 자동 분류 결과



(5) 중복 기사 제거

언론사별로 같은 기사를 중복으로 보낸 경우나 언론사 간 거의 비슷한 기사들을 분석 대상에서 제외하기 위해서 중복 기사 제거 작업을 실시하였다. 중복된 기사를 제거한 후 대표 기사 하나만을 남겨 놓았는데 그 기준은 완전히 중복된 기사 중 작성 시간이 가장 빠른 기사였다. 중복 기사 제거 작업은 클러스터링 작업과 연관되어 수행됐다. 1차년도에는 전체 문서에 대해서 자신을 제외한 나머지 전체 문서와의 코사인 유사도를 추출하여 유사도가 특정치 이상으로 높은 기사들에 대해서 중복 기사로 판별하고 제거하는 방식으로 진행했다. 이 방법의 경우 결과의 품질은 우수한 수준이었으나 전체 기사 수가 늘어날수록 연산량이 기하급수적으로 늘어나 시스템에 문제를 주는 경우가 많았다. 이러한 문제로 인해 전체 기사에 대해서 추출하지 않고 같은 날 발행한 기준으로만 중복 기사를 판단해 비교 기사 수를 줄이도록 했으나 날짜가 변경되면 중복된 기사로 판별하지 못하는 문제점도 가지고 있었다.

〈그림 3-10〉 중복 기사 예시

원본기사

×

| 언론사 | 날짜 | 제목 | News ID |
|------|----------|---|----------------------------|
| 한국경제 | 20160601 | [HEI포토] EXID LE, '독특한 뿔어진 청바지 입고' | 02100601.20160601142002491 |

기사의 중복기사 : 20개의 기사 (중복도 : 0.7)

| 언론사 | 날짜 | 제목 | News ID | 중복도 |
|------|----------|---|----------------------------|--------------------|
| 한국경제 | 20160601 | [HEI포토] EXID LE, '갈록한 허리 드러내고~' | 02100601.20160601142005273 | 1.0000000000000002 |
| 한국경제 | 20160601 | [HEI포토] EXID, '첫 정규앨범으로 돌아왔어요~' | 02100601.20160601115450925 | 0.9550125478392851 |
| 한국경제 | 20160601 | [HEI포토] EXID 경화, '사랑스러운 막내 미소~' | 02100601.20160601144122213 | 0.9392070991136933 |
| 한국경제 | 20160601 | [HEI포토] EXID 솔지, '깨물어주고 싶은 깜찍한 모습' | 02100601.20160601140912401 | 0.9214603932283949 |
| 한국경제 | 20160601 | [HEI포토] EXID 솔지, '사랑스럽게 손가락 하트~' | 02100601.20160601141959898 | 0.9214603932283949 |
| 한국경제 | 20160601 | [HEI포토] EXID 하니, '예고 가득한 미소~' | 02100601.20160601143037134 | 0.9190969416364235 |
| 한국경제 | 20160601 | [HEI포토] EXID 해린, '예고 가득한 미소~' | 02100601.20160601143044100 | 0.916564918850174 |
| 한국경제 | 20160601 | [HEI포토] EXID 해린, '깨물어주고 싶은 깜찍한 미소~' | 02100601.20160601143045776 | 0.9165649188501739 |
| 한국경제 | 20160601 | [HEI포토] EXID LE, '도도한 카리스마~' | 02100601.20160601121532181 | 0.8796937823415545 |
| 한국경제 | 20160601 | [HEI포토] EXID 경화, '기도하는 소녀처럼~' | 02100601.20160601121534682 | 0.8118484956630941 |

2차년도부터는 클러스터링 알고리즘으로 ‘DBSCAN’을 사용하면서 중복 기사 제거 작업도 이 알고리즘을 통해서 나온 결과물에서 유사도가 높은 기사를 제거하는 방식으로 진행했다. 코사인 유사도와 다르게 DBSCAN 알고리즘은 직관적으로 중복도를 판별할 값을 정하는 데 문제가 있었다. DBSCAN은 밀도 기반의 클러스터링으로 점이 세밀하게 몰려 있어서 밀도가 높은 부분을 한데 묶는 방식이다. 이때 밀도의 값을 임의로 지정을 해야 하는데 이 값을 엡실론(epsilon, eps) 값이라고 한다. 코사인 유사도의 경우 1에 가까울수록 유사도가 높다고 판단할 수 있었지만 eps 값은 그 정의가 없어 여러 값을 조정하면서 최종 중복 여부를 판단할 수밖에 없었다. 이에 따라 적절한 유사도 선정을 위해서 결과물들을 검수하면서 보정하는 작업을 반복적으로 수행했다.

(6) 클러스터링(주제 찾기 및 묶기)

중복 기사가 제거된 이후에는 클러스터링 작업을 실시했다. 클러스터링은 비슷한 주제의 기사들을 자동으로 묶어주는 알고리즘을 말한다. 수많은 기사 중 비슷한 기사들을 묶은 후 그중 위원회에서 정한 원칙에 따라 가장 대표성이 높은 기사를 선정하기 위한 목적이었다.

1차년도에는 클러스터링을 위해서 ‘k-means’ 알고리즘을 사용할 예정이었으나 기사 간에 코사인 유사도 측정을 통해 유사한 기사들을 묶어주는 방식으로 구현했다. 유사도가 너무 높으면 중복 기사, 너무 낮으면 상관없는 기사 등으로 분류했다. 여기서 적절한 유사도를 선별하는 작업은 계속 유사도를 변경하면서 적절한 유사도 값을 선택하는 방식으로 진행되었다.

〈그림 3-11〉 클러스터링 결과 화면 예시

| 뉴스 배열 체험하기 | |
|---|------------------|
| <p>김정은 訪中 성사 여부 최대 관심... 북핵 문제도 논의</p> <p>김정은 친서에 '핵문제 성의' 담았나... 북·중 대화국면 '촉각' - 한겨레</p> <p>북·중 관계개선 첫발 내디뎠지만... 문제는 '북핵' - 세계일보</p> <p>중외2가 "북·중 관계 발전 노선" 북에 신중 줄 가능성 낮아 - 매일경제</p> <p>北·中 "핵" 온도차 여전... 협력 북원 '채택 유력' '분수령' - 문화일보</p> <p>시진핑 '채택 관계개선' 두 톱까 잡을까 - 세계일보</p> | 총 36개의 기사가 있습니다. |
| <p>세누라-더민주, 국회의원 놓고 또 충돌... "여당이 해아" vs "정상적인 협상 안돼"</p> <p>더민주 "법사위원장 양보"... 황 건재" 운영에 확보 요청 - 경향신문</p> <p>"청와대 개입 아니냐"... 원세 비난에 감정싸움 치닫는 원구성 협상 - 서울경제</p> <p>다시 국회의원장 쟁탈전... 원구성 협상 '돌고 돌아 재차라' - 한겨레</p> <p>鼎野 "3각 관계"에 꼬이는 원구성... 비효율 현실화 - 문화일보</p> <p>여야, 국회의원장다 구성 놓고 정면 충돌 - 중도일보</p> | 총 35개의 기사가 있습니다. |
| <p>남양주 지하철 공사현장 붕괴 사고 4명 사망, 8명 부상</p> <p>지하서 용단작업 중 항... 지하철 공사장 붕괴 4명 사망 - 세계일보</p> <p>[남양주 지하철 공사장 붕괴, 4명 사망] 또 사람... 안전불감증 언제까지 - 서울경제</p> <p>또 사람... 남양주 지하철 공사장 가스폭발 14명 사상 - 매일경제</p> <p>[남양주 지하철공사장 폭발사고] 보름간 안전한국 출현... 진압에 '가스폭발' - 매일드림</p> <p>지하15m면 하강노동자만... 원형 직원은 없었다 - 한겨레</p> | 총 32개의 기사가 있습니다. |

2차년도에는 클러스터링 알고리즘들을 비교 분석하면서 1차년도에 적용하지 못했던 ‘k-means’ 알고리즘을 적용해서 클러스터링 작업을 진행할 계획이었다. ‘k-means’ 알고리즘은 클러스터링시 k값을 사전에 정의해야 하는 문제가 있었는데, 이보다 수행 속도와 정확도가 빠른 DBSCAN 알고리즘을 적용했다. DBSCAN(Density-based spatial clustering of applications with

noise) 알고리즘의 경우 사전 정의 없이 유사도만으로도 판단이 가능하며 수행 속도가 빠르고 정확도가 높다는 평가를 받았다. ‘k-means’나 계층적 군집 클러스터링의 경우 군집 간의 거리를 이용하여 클러스터링을 하는 방법인데, DBSCAN은 밀도 기반의 클러스터링으로 점이 세밀하게 몰려 있어서 밀도가 높은 부분을 클러스터링하는 방식이다. 어느 점을 기준으로 반경 x 내에 점이 n 개 이상 있으면 하나의 군집으로 인식한다. 코사인 유사도 방식과 마찬가지로 DBSCAN 알고리즘에서도 엡실론(epsilon, eps) 값 설정에 따라 클러스터링된 결과물이 변경되기 때문에 중복 기사 제거 작업과 마찬가지로 적절한 eps 값 선정을 위해서 결과물들을 검수하면서 보정하는 작업을 수행하였다.

2) 주요 계량 요인 11개의 추출 및 점수 부여 방식

뉴스 트러스트 위원회의 결정에 따른 계량 요인 개발은 약 30여 개 정도 이루어졌지만 실제 개발 과정에서 정확도 측면에서 활용 가능하다고 판단되는 요인은 제한적이었다. 예를 들어, 무주체 술어의 수, 익명 인용의 수 등은 추출을 위한 개발 작업은 이루어졌지만 정확도 측면에서 부족한 점이 많아 실제 활용하지는 못했다. 그 결과 기사 명, 기사의 길이, 인용문 수, 제목의 길이, 제목의 물음표 및 느낌표 수, 수치 인용 수, 이미지의 수, 평균 문장의 길이, 제목의 부사 수, 문장당 평균 부사 수, 기사 본문 중 인용문 비중 등 11개 요인만을 뉴스 트러스트 알고리즘의 요인으로 활용하였다.

(1) 기사 명

기사 명의 경우, 기자가 직접 쓴 기사는 믿을 만하다(trustworthy)고 판단할 수 있으며, 기명이 아닌 경우는 기사 작성과정에서 어뷰징 등 비정상적 요인의 개입이 있었다고 판단할 수 있다는 점을 감안하여, 뉴스 트러스트 위원회는 기사 본문 내 명기돼 있는 기자 이름을 최대한 추출했다. 추출방법은 기사 명 DB 필드에 기자 명이 있을 경우는 이를 통해 추출했고, DB 필드에는 없으나 본문에

기자 명이 있을 경우 언론사별 특성을 일일이 파악해 기계적으로 추출했다.

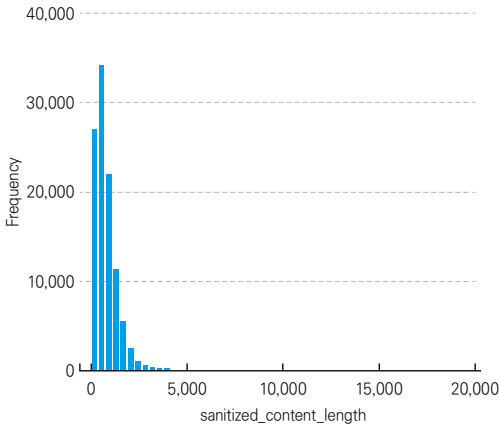
예를 들어, 언론사 코드 ‘01100101’은 본문에 기자 명이 잘 표기돼 있지 않고 수집 DB 필드에 잘 입력되어 있으며, ‘01100201’은 본문 하단의 불필요 정보를 제거하면 기자 명이 표기돼 있으며, 가끔 마지막 줄에 기자 명이 표기돼 있는 경우도 있었다. 이러한 방식으로 수집한 언론사들의 기자 명 표기 방식을 모두 정리하여 기자 명을 추출하였다. 그럼에도 불구하고 앞서 추출 방식에서 설명한 것처럼 각 언론사들마다 기자 명을 본문에 명기하는 방식이 상이하고 매번 같은 방식을 유지하는 것은 아니기에 100% 완벽하게 추출하지는 못했다. 추출된 기자 명들 중 인터넷뉴스팀 등 실명 기자 명이 아닌 경우 비실명 기자 명으로 별도 처리했다.

이후 기사 분류 및 매체 유형과 상관없이 실명 기자 명과 이메일이 병기돼 있을 경우에는 1점의 가점을 부여하고, 실명 기자 명만 있을 경우에는 0.8점, 비실명 기자 명일 경우에는 0점, 기자 명 없이 이메일만 있을 경우에는 0점을 부여했다. 기자 명 및 이메일이 아예 없을 경우에는 -1점의 감점을 부여했다. 이메일 명기 여부로 인해 점수 차등을 두는 것에 대해 위원들 간의 이견이 있었지만 저널리즘 관점에서 조금 더 바람직한 방향을 제시한다는 의미에서 이메일이 있을 경우 우대하는 것으로 결정됐다. 이메일을 명기하는 것은 기사의 투명성과 독자와의 소통 측면에서 의미가 있다는 위원회의 판단이었다.

(2) 기사의 길이

기사의 길이의 경우, 위원회는 길이가 길수록 사건과 관련한 완전한 내용(whole story)을 다루고 있을 가능성이 높으며, 정보량이 많고 심층성이 있으며, 다루는 정보의 범위가 넓어 좋은 기사라고 판단할 수 있다고 간주했다. 이에 따라 제목, 소재목, 사진 설명 등을 제외한 기사 본문의 길이라고 조작적으로 정의했다. 또한, 제목, 관련 기사, 광고 등 기사 본문 외 데이터를 정제하고 ‘utf-8’ 유니코드 기준으로 기사 본문을 원고지 글자 수로 길이를 계산했다.

[표 3-10] 기사의 길이 분포 통계



| 총 기사 수 | 105555 |
|-----------------|-----------|
| 평균 | 791.92 |
| 분산(Variance) | 436448.16 |
| 표준편차(Std. Dev.) | 660.64 |
| 왜도(Skewness) | 3.51 |
| 첨도(Kurtosis) | 36.93 |
| 최소값 | 0.00 |
| 최대값 | 18825.00 |
| 1% | 41.00 |
| 5% | 94.00 |
| 10% | 176.00 |
| 25% | 370.00 |
| 50% | 647.00 |
| 75% | 1045.00 |
| 90% | 1525.00 |
| 95% | 1895.00 |
| 99% | 3132.00 |

뉴스 트러스트위원회는 길이가 긴 것에 대한 보상이 있어야 한다는 관점에서 길이에 따라 0~1점 사이의 가점을 부여했다. 좀 더 구체적으로는 기사의 분류, 매체 유형(신문, 방송 등) 등에 따라 기사의 길이 평균을 추출한 후 평균 이상인 기사에 대해 표준편차에 따라 단계별 가점을 부여했다. 이에 따라 기사의 길이는 0, 0.165, 0.33, 0.495, 0.66, 0.835, 1 등 7단계의 값을 갖게 된다. 평균보다 기사의 길이가 길 경우 가점을 주고 너무 긴 기사가 너무 많은 점수를 가져가지 않도록 상한선을 설정한 것이며 매체와 분류에 따른 특성까지 고려해 점수를 부여한 것이다. 이러한 결정에는 [표 3-10]과 같은 통계 자료가 활용됐으며, 여기에 활용된 공식은 다음과 같다.

```

if (content_length < mean) then 0,
    else (content_length < mean + 0.5SD) then 0.165,
    else (content_length < mean + SD) then 0.33,
    else (content_length < mean + 1.5SD) then 0.495,
    else (content_length < mean + 2SD) then 0.66
    else (content_length < mean + 2.5SD) then 0.835
    else (content_length > mean + 2.5SD) then 1
    
```

(3) 인용문의 수

뉴스 트러스트 위원회는 기사 작성 시 기자가 다른 전문가 등의 발언을 인용하는 것은 기자의 사견이 개입하지 않음을 보여주기 때문에 저널리즘 관점에서 의미가 있다고 판단했다. 인용이 많을 경우 그만큼 기사를 신뢰할 수 있으며, 수용자들은 기사 내에 인용이 많을수록 그 기사의 내용을 믿을 만하다고 생각할 가능성이 높다.

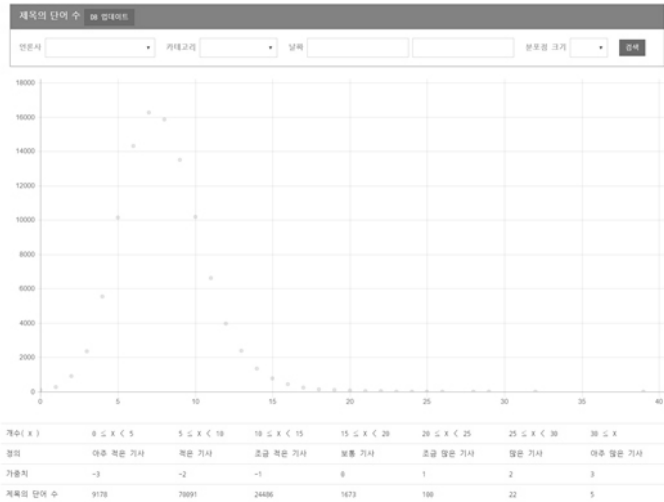
인용문은 다음과 같은 과정을 거쳐 추출됐다. 먼저, 기사 본문 내용 중 인용을 의미하는 큰따옴표(“)로 시작해서 큰따옴표(”)로 닫히는 내용들을 1차 추출한 후 큰따옴표 내에 형태소가 세 개 미만일 경우는 인용문이 아닌 강조문으로 판단하여 배제하고 세 개 이상인 경우만 인용문으로 추출해 그 수를 계산했다. 간접 인용의 경우는 추출 정확도가 떨어져 실제 반영하지는 않았다.

따라서 위원회는 기사 본문 내에 포함된 인용문의 개수가 많을수록 가점을 주기로 결정했다. 하지만, 인용문의 개수가 100개 이상인 기사가 있는 등 지나치게 많은 인용이 이루어지는 경우들이 발견됐다. 이에 따라 통계적 분포를 감안해 15개까지만 가점을 주는 방식으로 결정했다. 즉, 인용문이 0개일 경우는 0점, 1개일 경우는 1/15점이며, 15개 이상일 경우는 모두 1점을 부여했다.

(4) 제목의 길이

기사의 제목은 기사의 중요성과 내용을 판단할 수 있는 핵심적 정보라는 점에서 중요하다. 하지만, 이용자의 관심을 끌기 위해 제목에 많은 내용을 담으려는 시도들이 이어지고 있다. 뉴스 트러스트 위원회는 기사의 제목이 지나치게 길 경우 기사의 핵심 내용을 파악할 수 없게 만들어 기사 가치를 훼손할 수 있다고 판단했다. 이에 따라 ‘utf-8’ 유니코드 기준으로 기사의 제목을 원고지 글자 수로 길이 계산한 후 일정 길이보다 길 경우 감점을 부여하는 것으로 결정했다.

〈그림 3-12〉 기사 제목에 사용된 단어 수 분포도



〈그림 3-12〉과 같이 제목의 길이 통계 분포도를 검토한 결과, 지나치게 긴 제목에 대해서만 감점을 부여했다. 그 기준은 45자였는데 제목의 길이가 45자를 초과하는 기사의 비율은 5% 미만이었기 때문이다.

※ 제목이 45자 넘는 사례

- ‘또 오해영’ 에릭-서현진 비하인드컷...“아무리 망했어도!! 어떻게 그렇게 보내냐, 어떻게 그렇게 말하냐”
- [HET포토] 레이싱모델 문가경, ‘치맛자락 사이로 우월한 각선미 과시’ (2016 부산모터쇼)
- 2016 백상예술대상 유아인, 수상소감에 네티즌...“떨림이 느껴질 정도로 깊이 있는 수상소감 하는 건 외운 게 아니라 진심이다!”

(5) 제목의 물음표, 느낌표 수

기사의 제목에 물음표, 느낌표 등을 사용하는 것은 기사의 핵심내용을 전달하는 데 오히려 방해가 된다. 기사 제목에 물음표, 느낌표 등을 사용할 경우 기사 내용에 대한 선입견을 부여하고 내용을 과장할 가능성이 있기 때문이다.

뉴스 트러스트 위원회는 이에 따라 제목에 물음표, 느낌표 수가 일정 개수 이상일 경우 감점을 부여하기로 결정했다. 따옴표는 문제 삼지 않으며, 물음표, 느낌표를 사용했을 경우 사용 여부에 따라 감점을 부여하는데, 물음표와 느낌표의 사용한 개수가 1개인 경우는 0.5점을 감점하고, 2개 이상인 경우는 1점을 감점하였다. 제목에 물음표와 느낌표를 1개 이상 사용한 기사의 수가 전체 5% 미만이었기 때문이다.

(6) 수치 인용 수

기사 본문 내에 수치 등 객관적 정보가 많을수록 보다 믿을 수 있는 정보를 제공하고 있다고 볼 수 있다.

〈그림 3-13〉 기사 내 수치 인용 수 분포



뉴스 트러스트 위원회에서는 객관적인 수치 인용이 기사 내에 많을수록 정확한 정보를 제공하고 있다고 가정하였다. 기사 제목, 관련 기사, 광고 등 기사 본문 외 데이터를 제외하고 본문 내 수치가 많이 포함될수록 가점을 주는 것으로 합의하였다. ‘utf-8’ 유니코드 기준으로 기사 본문을 원고지 글자 수로 길이

계산 후 한글을 빼고 무조건 숫자만 추출하였다.

〈그림 3-13〉와 같이 기사 분류 및 언론사 유형에 따라 수치 인용이 다를 수 있음을 통계적으로 확인하였고, 기사의 길이와 유사한 방식으로 가점을 부여하기로 결정했다. 같은 분류, 같은 언론사 유형을 기준으로 평균 이하는 0점, 평균보다 0.5 표준편차 이하는 0.33점 등으로 차등하여 가점을 부여하였고 평균보다 1 표준편차 이상 많으면 모두 1점의 가점을 부여하였다. 단, 수치를 1, 2, 3 등과 같은 아라비아 숫자가 아닌 한글(일, 이, 하나, 둘 등) 혹은 영어(원, 투 등) 등 언어로 표현한 경우는 수치로 인용한 것으로 포함하지 못했다. 수치 인용 수에 활용된 공식은 다음과 같다.

```
if (content_numNumber < mean) then 0,  
    else (content_numNumber < mean + 0.5SD) then 0.33,  
    else (content_numNumber < mean + SD) then 0.66,  
    else 1
```

(7) 기사 내 이미지의 수

기사 내에 텍스트 외 이미지 등 멀티미디어 정보를 제공할 경우 독창성, 사실성, 심층성, 유용성이 높다고 가정할 수 있지만, 이미지의 수가 너무 지나칠 경우 기사 독이성에 저해된다고 뉴스 트러스트 논의에서 의견이 모아졌다. 뉴스 트러스트에서 활용한 빅카인즈 DB에는 동영상 정보가 누락되어 있어 기사 내에 삽입된 이미지의 수만 산출하였다. 기사 본문 내에 포함된 이미지 숫자의 총합을 계산하였고, 지나치게 많은 이미지는 오히려 가독성을 저해할 수 있다고 판단하여 3개를 기준으로 전후 차등하여 점수를 부여하는 것으로 합의됐다.

분석 대상 전체 기사를 대상으로 평균 이미지의 수를 계산한 결과 1.01개였기 때문이다. 이에 따라 기사 본문 내 이미지 개수가 0개는 0점, 1개 0.33, 2개 0.66, 3개 1점, 4개 0.66점, 5개 0.33점 6개 이상 0점을 부여하였다. 카드뉴스 등 이미지로만 구성된 기사의 경우에는 아예 점수를 부여하지 않았다.

(8) 평균 문장의 길이

기사는 핵심 내용을 압축적으로 전달해야 한다. 또한, 기사는 명확한 정보 전달을 목적으로 하고 있어 가능한 명료하게 작성해야 한다. 문장의 길이가 길면 독자의 가독성이 떨어지며, 명확한 의미 전달이 어려워질 수 있기 때문이다. 뉴스 트러스트 위원회는 이에 따라 기사 본문 내 문장들의 평균 길이가 길 경우 감점을 부여하기로 결정했다.

전체 기사의 길이를 기사 내 전체 문장 수로 나눠 전체 기사의 평균 문장의 길이보다 지나치게 길 경우 감점을 부여했다. 평균 문장의 길이보다 1 표준편차 이상 길 경우에는 감점 1점을 부여하고 그 외에는 모두 0점을 부여하였다.

(9) 제목의 부사 수

기사는 가능한 정보를 명확하게 전달해야 하는데, 주어, 동사, 목적어 이외에 형용사, 부사 등 수식어가 많을수록 정보 전달의 명료성이 떨어지며, 이용자들의 기사에 대한 신뢰도도 낮아질 수 있다. 특히, 기사 제목의 경우 명료성이 중요하기 때문에 부사 등 수식어 사용은 가급적 지양해야 한다. 뉴스 트러스트 위원회는 이에 따라 기사 제목 내에 사용된 부사의 수가 많을 경우에는 감점을 부여하기로 결정했다. 기사 제목의 형태소를 분석한 후 용언 또는 다른 말 앞에 놓여 뜻을 분명하게 하는 품사로서 부사의 개수를 산출했다. 분석 대상 기사들의 평균 제목의 기사 수는 0.18개였으며 분류 및 매체 유형별로 큰 차이는 없었다. 이러한 통계 수치를 감안하여 제목에 사용된 부사 수가 0개에서 1개까지는 감점하지 않고, 2개를 사용했을 경우에는 0.5점 감점, 3개 이상 사용했을 경우에는 1점을 감점하기로 합의했다.

(10) 문장당 평균 부사 수

정확한 정보 전달을 위해 기사는 가능한 명료해야 하기 때문에 기사의 제목뿐만 아니라 기사 내에서도 부사 등 수식어의 사용을 가능한 지양하는 것이 좋다. 주어,

동사, 목적어 이외에 형용사, 부사 등 수식어가 많을수록 정보 전달의 명료성이 떨어지며, 이용자들의 기사에 대한 신뢰도도 낮아질 수 있기 때문이다. 이에 따라 뉴스 트러스트 위원회는 전체 기사 내에 부사의 사용이 많으면, 독이성과 사실성이 떨어지고 선정성이 높아질 수 있다고 판단하고, 형용사, 접속사는 제외하고 전체 기사 내에서 부사가 지나치게 많이 사용되면 감점을 부여하기로 결정했다.

기사 본문에 대한 형태소 분석을 통해 사용된 부사의 수를 추출하고 이를 기사 본문의 문장 수로 나눠 1개 문장당 평균 부사의 수를 산출했다. 지나치게 많은 것에만 감점을 부여하기로 결정하고 평균 문장당 부사 수보다 2 표준편차보다 많은 경우에만 1점을 감점하기로 했다. 또한, 통계적 검토 결과 기사의 분류와 매체 유형에 따라 평균 개수가 상이함을 발견하고 분류 및 매체 유형에 따라 별도로 적용하기로 결정했다.

(11) 기사 본문 중 인용문의 비중

기사 작성 시 관련된 전문가 등의 인용을 하는 것은 기자의 사견이 개입하지 않음을 보여주지만, 지나친 인용은 사실의 확인 없이 일방적인 주장만을 전달하고 있을 가능성이 있다. 특히, SNS 발언 내용, TV와 라디오 방송 내용 등 취재 없이 일방적 전달 기사의 경우 지나친 인용이 문제가 되는 사례가 다수 존재한다. 이에 따라 뉴스 트러스트 위원회는 기사 내 인용문의 비중이 지나치게 높을 경우에는 감점을 부여하기로 결정했다. 기사 본문 내용 중 인용을 의미하는 큰따옴표(“)로 시작해서 큰따옴표(”)로 닫히는 내용들을 1차 추출한 뒤 큰따옴표 내에 형태소가 세 개 미만일 경우는 인용문이 아닌 강조문으로 판단하여 배제하고 세 개 이상인 경우만 계산하였다. 추출된 인용문의 길이를 산출한 뒤 전체 인용문의 길이를 전체 기사의 길이로 나눴다. 분석 대상 기사의 평균 인용문 비중은 0.15였는데, 정치 기사의 경우가 0.2로 가장 높았으며 연애가 0.16으로 다음이었다. 이러한 통계치를 감안해 뉴스 트러스트 위원회는 전체 기사에서 인용문의 비중이 0.5~0.8까지는 0.5점을 감점하고 0.8 이상은 1점을 감점하는 것으로 결정했다.

3) 기사별 점수 부여 방식과 배열

11가지 요인별로 추출이 끝나면 [표 3-11]과 같은 각각의 점수 부여 방식에 따라 개별 기사들은 기본 점수를 부여받게 된다. 각 요인에 따라 가점을 받거나 감점을 받은 값을 모든 더한 값이 각 기사가 갖게 되는 기본 값이다. 예를 들어, 기자 명이 있는데 이메일이 없는 기사는 기자 명 요인에서 0.8점의 가점을 받고, 기사의 길이가 분류 및 매체 유형 평균 더하기 1 표준편차보다 크고 1.5 표준편차보다 작다면 0.495점의 가점을 받는다. 하지만, 기사의 제목에 부사가 2개 사용됐다면 0.5점의 감점을 받게 된다. 나머지 요인들에서 가점이나 감점을 받지 않았다면 해당 기사의 계량 요인에 따른 기본 값은 0.795점이 된다. (공식 1)

이렇게 계산된 값은 각 요인이 뉴스 트러스트 위원회가 설정한 저널리즘 요인과의 관계에 따라 저널리즘 가치 값이 매겨진다. [표 3-9]에서 제시한 바와 같이 저널리즘 가치와 계량 요인 간의 관계에 따라 가중치가 0일 경우에는 해당 요인이 특정 저널리즘 가치와 전혀 관련이 없는 것으로 간주된다. 예를 들어, 기자 명 요인의 경우는 균형성, 다양성, 독이성, 독창성, 사실성, 심층성, 유용성, 중요성, 투명성, 선정성 가치와 모두 관련을 갖고 있지만, 제목의 부사 수는 균형성, 독이성, 사실성, 선정성과만 관련이 있다. 관련이 있을 경우에는 해당 저널리즘 가치별로 각 요인에 따른 기본 점수를 받게 되지만 관련이 없을 경우에는 점수를 받지 못한다. 제목의 부사 수에서 감점이 됐다면 균형성, 독이성, 사실성, 선정성 가치에서는 0.5점이 감점되지만 나머지 저널리즘 가치에서는 감점이 되지 않는다. 이렇듯 각 저널리즘 가치별로 점수를 부여받게 되고 이를 더한 값이 저널리즘 가치 점수가 된다. (공식 2)

[표 3-11] 11개 계량 요인별 기본 점수 부여 방식

| 계량 요인 | 점수 부여 방식 |
|-----------------|--|
| 기자 명 | <ul style="list-style-type: none"> • 기자 명 DB 필드를 통해 추출 • DB 필드에는 없으나 본문에 기자 명이 있을 경우 언론사별 패턴 분석을 통해 기계적으로 추출 • 추출된 기자 명들 중 인터넷뉴스팀 등 기자가 작성하지 않은 특수 케이스 처리 • 실명 기자 명과 이메일 있을 경우 1점, 실명 기자 명만 있으면 0.8점, 기자 명이 아예 없으면 -1점, 비실명 기자 명만 있으면 0점 |
| 기사의 길이 | <ul style="list-style-type: none"> • 길이에 따라 0~1점 사이 가점 부여(길이가 긴 것에 대한 보상이 있어야 함) • 평균 이상인 기사에 대해 단계별 가점(표준편차 따른 단계 구분), 분류(category) 및 신문·방송 등 유형에 따른 적용 필요 • if (content_length < mean) then 0, else (content_length < mean + 0.5SD) then 0.165, else (content_length < mean + SD) then 0.33, else (content_length < mean + 1.5SD) then 0.495, else (content_length < mean + 2SD) then 0.66, else (content_length < mean + 2.5SD) then 0.835, else (content_length > mean + 2.5SD) then 1 |
| 인용문의 수 | <ul style="list-style-type: none"> • 0~15개까지 균등하게 점증 가점 • 15개 이상은 모두 1점 |
| 제목의 길이 | <ul style="list-style-type: none"> • 명확한 기준을 세우기 애매하기 때문에 지나치게 긴 제목에 대해서만 감점 • 제목의 길이가 45자를 넘으면 감점 |
| 제목의 물음표, 느낌표 수 | <ul style="list-style-type: none"> • 따옴표는 문제 삼지 않으며, 물음표, 느낌표를 사용했을 경우 사용 여부에 따라 감점 • 물음표 느낌표 1개인 경우 -0.5, 2개인 경우 -1점 |
| 수치 인용 수 | <ul style="list-style-type: none"> • 분류 및 유형 별로 숫자가 많으면 가점, 한글을 빼고 무조건 숫자만 추출 • 평균 이하는 0점, 0~0.5SD=0.33, 0.5SD~1SD=0.66, 1SD 이상 = 1 • 기사 분류 및 매체 유형에 따라 평균 및 표준편차 적용 |
| 이미지의 수 | <ul style="list-style-type: none"> • 3개를 기점으로 전후 차등하여 점수 부여 • 0개 = 0점, 1개 = 0.33, 2개 = 0.66, 3개 = 1점, 4개 = 0.66점, 5개 = 0.33점, 6개 이상 = 0점 |
| 평균 문장의 길이 | <ul style="list-style-type: none"> • 문장의 길이가 평균적으로 지나치게 길 경우 감점 • 평균 + 1SD 이상인 경우 -1점, 나머지는 0점 • 분류 및 매체 유형에 따라 평균 및 표준편차 적용 |
| 제목에 사용된 부사 수 | <ul style="list-style-type: none"> • 부사 수가 많을 경우 감점(제목에 사용된 경우로 개수별로 감점) • 0에서 1까지는 0점, 2개는 -0.5점, 3개 이상은 -1점 |
| 문장당 평균 부사 수 | <ul style="list-style-type: none"> • 형용사, 접속사는 제외하고 부사가 지나치게 많은 것에만 감점(분류 및 매체 유형에 따라 적용) • 평균값 + 2SD보다 많은 경우 -1점 적용 |
| 기사 본문 중 인용문의 비중 | <ul style="list-style-type: none"> • 전체 기사 내에서 인용문이 차지하는 비중이 지나치게 높을 경우 감점 • 전체 기사에서 인용문의 비중이 0.5~0.8까지는 -0.5, 0.8이상은 -1점 |

이후 [표 3-9]의 기준에 따라 저널리즘 가치별 가중치를 곱한다. 예를 들어, 제목의 부사 수와 관련해 0.5점 감점을 받았을 경우, 균형성과 관련해서는 0.5 (기본 점수)에 0.501(가중치)를 곱한 0.2505점을 감점 받게 된다. 독이성과

관련해서는 0.5(기본 점수)에 2.466(가중치)을 곱한 1,233점을 감점 받게 된다. 이렇듯 각 저널리즘 가치별로 가중치를 곱해서 최종적으로 얻게 된 값을 모두 합친 값이 저널리즘 가중치 점수가 된다. (공식 3)

이 과정을 기사 하나를 예시로 들어 설명한다.

“코로나19, 악의적인 ‘교란정보’가 가장 큰 문제”

[미디어오늘 정철운 기자]

감염병은 잘못된 정보가 바이러스처럼 퍼지는 것이 실제 바이러스 확산만큼 문제다. 이는 한국만의 문제도 아니다. ‘미디어와 코로나 판데믹(pandemic, 전염병의 지구적 유행)’을 주제로 한국언론진흥재단(이사장 민병욱)이 9일 개최한 한미 언론 합동 토론회에서는 ‘치료적 보도’로서 언론의 역할에 대한 주문과 함께 다양한 제언이 등장했다.

코로나19 국면에서 언론의 점수는 좋지 않다. KBS 공영미디어연구소가 3월 31일~4월 2일까지 성인 남녀 1,069명을 대상으로 진행한 설문 결과 주요 대응 주체별 평가에서 질병관리본부(91.6%)가 가장 잘 대응하고 있다는 평가를 받은 반면, 언론·뉴스미디어가 잘 대응했다는 평가는 46%에 그쳐 6개 대응 주체 중 최하위를 나타냈다. 같은 조사에서 정부는 67.2%, 대통령 및 청와대는 64.5%로 나타났다. 또한 응답자의 56%는 국내 언론이 ‘검증되지 않은 정보와 가짜 뉴스를 많이 전달하고 있다’는 주장에 동의했다.

한국언론진흥재단 미디어연구센터가 3월 9일부터 12일까지 성인 1,000명을 대상으로 실시한 온라인 설문조사에도 코로나19와 관련해 사회의 여러 주체가 적절히 대응하고 있는지를 4점 척도로 물은 결과에서도 언론(63.7%)은 의료기관(93.2%), 지방자치단체(77.2%), 정부(74.4%)보다 낮은 신뢰도를 나타냈다. 확진자 또는 유증상자(54.1%) 다음으로 신뢰하지 못하는 주체가 언론이었다.

코로나19 국면에서 언론의 이 같은 낮은 신뢰도와 관련, 이날 발제를 맡은 정준희 한양대 정보사회미디어학과 겸임교수는 ‘교란정보’라는 개념을 사용해 설명했다. 정준희 교수는 “문제해결을 촉진시킨다는 면에서 미디어가 주는 공포의 순기능이 있지만 자극의 정도를 높여가는 역기능도 만만치 않았다.”고 지적하며 “‘똥렸다’, ‘창궐’, ‘대혼란’ 같은 식의 과도한 기술을 통해 지속적으로 자극을 높여가며 각자가 이기적 선택을 하도록 조장한 면이 있는지 반성해야 한다.”고 했다. 정 교수는 “뉴스가 감염병 국면에 도움이 못 되고 오히려 그릇된 감정을 전염시키는 경우가 있다.”며 지금 같은 인포데믹

(infodemic, 잘못된 정보가 전염병처럼 퍼지는 현상) 국면에선 치료적 보도가 필요하다고 지적한 뒤 “악의적이고 의도적으로 유포되는 정보를 의미하는 ‘교란정보(DISINFORMATION)’가 가장 큰 문제”라며 “교란정보에 맞춘 정보제공이 중요하고 이는 기성 언론을 통해 이뤄지는 것이 바람직하다.”며 ‘교정정보’를 강조했다.

우리는 교란정보를 극복하고 있을까. 정준희 교수는 “현재 해외의 상황 변화에 따라 교정정보가 상당 부분 유통됐다. 해외 상황이 객관적으로 드러나면서 한국과 비교한 교정정보가 나타났다.”고 지적했다. 예컨대 일부 국내 언론이 지속적으로 방역 당국에 대한 ‘악의적’ 보도를 이어갔으나, “바이러스가 시민적 자유를 시험하는 시대에, 도시를 계속 열어두면서 감염을 공격적으로 감시하는 이 전략이 먹히기만 한다면 민주사회에 본보기가 될 수 있다.”(2/25)고 보도한 뉴욕타임스, “WHO(세계보건기구)는 한국이 코로나바이러스 감염 제한에 성공한 점을 들어 모든 나라에게 공격적인 테스트를 촉구했다.”(3/18)고 보도한 가디언 같은 외신보도가 국내에 확산되면서 국내 방역 당국의 수준이 객관적으로 드러났고, 이것이 ‘교정정보’ 역할을 했다는 설명이다. 정준희 교수는 “외신을 인용하는 언론 보도가 증가하며 팩트체크 노력이 이어졌고, 여기에 한국 방역 당국의 투명한 정보공개와 교정정보 확산 노력이 더해졌다.”고 밝혔다.

이 같은 결과 방역 당국을 비판하는 언론보다 방역 당국을 신뢰하는 여론이 나타나고 있다는 지적이다. 조슈아 벤트 하버드대 니먼저널리즘랩 소장은 “어느 때보다 뉴스 이용자가 증가하고 있지만 언론사는 재정적인 어려움에 직면했다.”고 지적했다. 조슈아 벤트 소장은 “코로나19는 현지의 지역 뉴스가 가장 중요하지만 대부분 재정적 어려움을 겪으며 문을 닫고 있다. 언론사들이 준비했던 각종 콘퍼런스는 취소되고 있고, 신문 배달도 위기 상황이다. 모든 매체의 광고매출이 감소하고 있고 언론사들은 비용축소와 통폐합이 예상된다.”며 저널리즘의 붕괴를 우려했다.

이는 국내 언론이 처한 상황과도 유사하다. 에이미 브리튼 워싱턴포스트 기자는 “이런 위기상황에는 검증 시간이 충분하지 않다. 대면 취재도 어려운 상황이다. 병원 간의 정보공유도 부족하다. 광고 수입도 급감했다. 그럼에도 불구하고 최근 워싱턴포스트의 구독자가 어느 때보다 오르고 있다. 독자들은 진짜 정보를 원하고 있다.”며 전 세계 기자들의 분투를 당부했다.

정철운 기자

위 기사는 공식 1에 의거해 [표 3-12]와 같은 기본 점수 값을 부여받게 된다.

[표 3-12] 공식 1에 따른 예시 기사의 기본 점수

| 계량 요인 | 추출 내용 | 기본 점수 |
|---------------|----------------|--------------|
| 바이라인 | name: "정철은 기자" | 0.8 |
| 이미지 개수 | 0개 | 0 |
| 기사 길이 | 2,251 | 1 |
| 평균 문장 길이 | 96.913 | 0 |
| 문장당 평균 부사 수 | 0개 | 0 |
| 제목 길이 | 29 | 0 |
| 제목에 물음표·느낌표 수 | 0개 | 0 |
| 제목의 부사 수 | 1개 | -0.5 |
| 수치 인용 수 | 38개 | 1 |
| 인용문 수 | 14개 | 0.933 |
| 인용문 길이 비율 | 43.22% | 0 |
| 총점 | | 3.233 |

[표 3-12]의 기본 점수는 공식 2에 따라 [표 3-13]과 같은 저널리즘 가치 점수 값을 갖게 된다.

이렇듯 기본 점수 3.233점을 받았던 예시 기사는 저널리즘 가치별로 점수를 부여받은 후 이들을 각각 더해 31.4점의 저널리즘 가치 점수를 갖게 된다. 10개 (반복성은 실제 적용에서 제외)의 저널리즘 가치 점수들과 계량 요인과의 관계에 따라 값이 달리 적용되고 이를 모두 더하게 돼 기본 점수보다 높은 값을 갖게 된 것이다. 예시 기사의 경우 가점 요인이 많아 기본 점수보다 저널리즘 가치 점수가 높았지만, 감점 요인이 많았을 경우에는 저널리즘 가치 점수가 기본 점수보다 크게 낮아질 수도 있다.

[표 3-13] 공식 2에 따른 예시 기사의 저널리즘 가치 점수

| 저널리즘 가치 | 저널리즘 가치 점수 |
|---------|--|
| 독이성 | 기자 명(0.8) + 기사 길이(1) + 인용문 수(0.933) + 제목 길이(0) + 제목 물음표·느낌표(0) + 수치 인용 수(1) + 이미지 수(0) + 평균 문장 길이(0) + 제목 부사 수(-0.5) + 문장 평균 부사 수(0) = 3.233 |
| 투명성 | 기자 명(0.8) + 기사 길이(1) + 인용문 수(0.933) + 제목 물음표·느낌표(0) + 수치 인용 수(1) + 이미지 수(0) + 인용문 비중(0) = 3.733 |
| 사실성 | 기자 명(0.8) + 기사 길이(1) + 인용문 수(0.933) + 제목 물음표·느낌표(0) + 수치 인용 수(1) + 이미지 수(0) + 제목 부사 수(-0.5) + 문장 평균 부사 수(0) + 인용문 비중(0) = 3.233 |
| 유용성 | 기자 명(0.8) + 기사 길이(1) + 인용문 수(0.933) + 수치 인용 수(1) + 이미지 수(0) = 3.733 |
| 균형성 | 기자 명(0.8) + 기사 길이(1) + 인용문 수(0.933) + 제목 물음표·느낌표(0) + 제목 부사 수(-0.5) + 문장 평균 부사 수(0) + 인용문 비중(0) = 2.233 |
| 다양성 | 기자 명(0.8) + 기사 길이(1) + 인용문 수(0.933) + 제목 길이(0) + 수치 인용수(1) + 이미지 수(0) + 평균 문장 길이(0) + 인용문 비중(0) = 3.733 |
| 독창성 | 기자 명(0.8) + 기사 길이(1) + 인용문 수(0.933) + 제목 물음표·느낌표(0) + 수치 인용수(1) + 이미지 수(0) = 3.733 |
| 중요성 | 기자 명(0.8) + 기사 길이(1) + 인용문 수(0.933) + 수치 인용 수(1) + 이미지 수(0) = 3.733 |
| 심층성 | 기자 명(0.8) + 기사 길이(1) + 인용문 수(0.933) + 수치 인용 수(1) + 이미지 수(0) + 인용문 비중(0) = 3.733 |
| 선정성 | 기자 명(0.8) + 제목 길이(0) + 제목 물음표·느낌표(0) + 제목 부사 수(-0.5) + 문장 평균 부사 수(0) + 인용문 비중(0) = 0.3 |
| 총합 | 31.4 |

[표 3-13]의 저널리즘 가치 점수는 공식 3에 따라 [표 3-14]와 같은 가중치 점수를 갖게 된다. 각 저널리즘 가치에 따른 요인별 가중치를 곱한 값을 더한 점수다.

저널리즘 가치 점수 31.4점을 받았던 예시 기사는 저널리즘 가치별 가중치가 곱해지면서 총합 85.44점의 점수를 받게 됐다. 3.233점이었던 기본 점수가 저널리즘 가치 점수를 부여하고 가치에 따른 가중치를 곱하면서 25배 가량 높은 점수를 갖게 된 것이다. 뉴스 트러스트 알고리즘의 핵심이다. 뉴스 트러스트 알고리즘은 단순히 형식적 계량 요인을 추출해 점수를 부여하는 것을 넘어 각 요인이 저널리즘 가치와 갖는 연관성을 분석하고, 그 분석에 따라 가중치를 더해 최종적인 배열 기준으로 활용하는 것이다.

[표 3-14] 공식 3에 따른 예시 기사의 저널리즘 가중치 점수

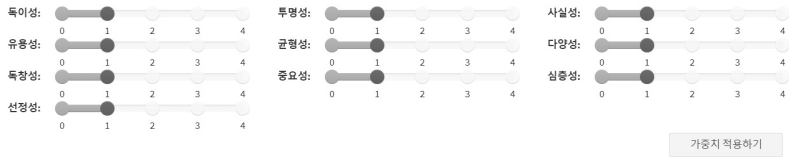
| 저널리즘 가치 | 저널리즘 가중치 점수 |
|---------|--|
| 독이성 | $0.8 \times 0.001 + 1 \times 0.003 + 0.933 \times 0.001 + 0 \times 1 + 0 \times 1.002 + 1 \times 1.354 + 0 \times 1.5 + 0 \times 1.5 + -0.5 \times 2.466 + 0 \times 0.5 = 0.125$ |
| 투명성 | $0.8 \times 4.498 + 1 \times 3.003 + 0.933 \times 4.5 + 0 \times 3.619 + 1 \times 1.454 + 0 \times 1 + 0 \times 0.001 = 12.255$ |
| 사실성 | $0.8 \times 4.493 + 1 \times 3.503 + 0.933 \times 3.501 + 0 \times 0.001 + 1 \times 0.502 + 0 \times 1.5 + -0.5 \times 0.5 + 0 \times 1.5 + 0 \times 1 = 10.617$ |
| 유용성 | $0.8 \times 3.494 + 1 \times 3.498 + 0.933 \times 2.001 + 1 \times 1.956 + 0 \times 1 = 10.116$ |
| 균형성 | $0.8 \times 2.996 + 1 \times 3.002 + 0.933 \times 3 + 0 \times 1.501 + -0.5 \times 0.501 + 0 \times 1 + 0 \times 1 = 7.948$ |
| 다양성 | $0.8 \times 0.998 + 1 \times 4.994 + 0.933 \times 2.501 + 0 \times 0.5 + 1 \times 1.953 + 0 \times 1 + 0 \times 0.5 + 0 \times 0.5 = 10.079$ |
| 독창성 | $0.8 \times 4.494 + 1 \times 4.492 + 0.933 \times 3.501 + 0 \times 3.09 + 1 \times 1.823 + 0 \times 1.501 = 13.177$ |
| 중요성 | $0.8 \times 2.495 + 1 \times 3.503 + 0.933 \times 3.5 + 1 \times 1.002 + 0 \times 0.5 = 9.767$ |
| 심층성 | $0.8 \times 4.496 + 1 \times 4.995 + 0.933 \times 3.501 + 1 \times 1.336 + 0 \times 1 + 0 \times 1 = 13.195$ |
| 선정성 | $0.8 \times 4.491 + 0 \times 3.5 + 0 \times 3.501 + -0.5 \times 3.5 + 0 \times 3.5 + 0 \times 3.5 = 1.842$ |
| 총합 | 85.44 |

앞서 밝혔듯이 뉴스 트러스트 알고리즘은 모든 기사들을 점수에 따라 서열화하는 것이 목적이 아니다. 계량적인 기사 배열에 있어서 아직 완전한 답은 없기 때문이다. 따라서 이용자로 하여금 저널리즘 가중치를 스스로 설정하여 배열 순서를 바꿔볼 수 있도록 하였다. 균형성, 심층성 등을 더 중요하게 간주하거나 다양성, 독창성 등의 비중을 낮출 수 있도록 한 것이다. <그림 3-14>과 같이 이용자들은 본인이 직접 저널리즘 가중치를 0~4까지 5단계로 설정해 볼 수 있다. 이를 통해 이용자들은 자신들이 중요하게 생각하는 저널리즘 가치에 따라 기사의 배열이 어떻게 달라지는지를 볼 수 있다. 단, 이러한 배열은 비슷한 주제로 묶인 클러스터 내 기사들만을 대상으로 한다. 지속적으로 강조하듯이 뉴스 트러스트 알고리즘은 전체 기사를 대상으로 서열화하는 것이 목적이 아니기 때문이다.

〈그림 3-14〉 이용자의 요인별 저널리즘 가중치 설정 화면

요인별 가중치 적용 패널

*[요인별 가중치 적용 패널]로 적용한 점수는 본 화면에서만 확인되며, 기사를 클릭해 보는 상세계산에는 반영되지 않습니다.



뉴스 트러스트 알고리즘에서 활용한 11개 계량 요인은 뉴스 트러스트 위원회가 목표로 한 평가 요인들 중 자체적으로 평가했을 때 추출 정확도가 90% 이상인 것만을 적용한 것이다. 완성품이 아닌 중간 단계의 결과물이다. 뉴스 트러스트 위원회는 이러한 한계로 인해 모든 개발 결과물을 소스코드 형식으로 깃허브에 공개하였다. 공개된 주소¹¹⁾에 접속하면 개발 내용을 모두 확인할 수 있는데, 이는 소스코드 공개를 통해 언론사, 일반 이용자 등의 다양한 의견을 수렴하고 보다 나은 적용 방식을 찾기 위한 방안이었다.

11) <https://github.com/KPF-NEWSTRUST>

4. 뉴스 트러스트 알고리즘의 한계와 의미

뉴스 트러스트 알고리즘은 사실 한계가 명확하다. 뉴스 기사의 신뢰도를 측정하기 위해 활용한 11개 요인들은 모두 기존 자동 배열 사례들처럼 기사의 형식적 측면을 측정한 것이지, 의견의 다양성, 공정성 등 저널리즘의 질적 가치를 직접 측정해내지는 못했다. 또한, 각 요인들의 점수화 과정보다 자의적이라는 비판을 피할 수 없다. 뉴스 기사의 신뢰도를 측정한다고 했지만 신뢰도를 구성하는 극히 일부분만을 기계적으로 측정해 구현한 것에 지나지 않기 때문이다.

알고리즘의 영향력이 갈수록 확대되고 사회적 감시 필요성이 높아지고 있는 현실에서 미래 언론의 공공성 구현을 위한 선구적 의미를 지니고 해당 사업이 출범했지만, 애초의 목표대로 진행되지 못하고 초기 단계에서 사업을 종료하게 되었다는 점도 한계로 지적할 수 있다. 지금까지의 개발 내용은 플랫폼 기업들에서 적용되고 있는 기술 수준에 비하면 낮은 수준이기 때문이다. 초기 개발단계에서 중단된 뉴스 트러스트 알고리즘을 고도화해서 발전시키는 것이 필요하지만 현재 상태에서 뉴스 트러스트 알고리즘은 뉴스 판단의 실제적 기준의 잣대로 쓰이기는 어렵다.

신뢰할 수 있는 좋은 뉴스를 선택할 수 있는 조건에 대해 고민하면서 저널리즘을 연구하는 언론학, 커뮤니케이션학계의 경험과 텍스트와 관련한 알고리즘을 개발하는 컴퓨터 소프트웨어 전산학, 언어학 등의 전문성이 합쳐지면서 학문 간의 융합을 의도하였지만, 실제로는 공학적 접근보다는 이론적, 실무적 접근이 주로 이루어진 점도 한계로 지적할 수 있다. 신뢰라는 가치를 공학적, 수학적으로 규정하거나 측정하기 어려웠기 때문에 여러 문제들을 계량적인 방법론보다는 가치 지향적 접근을 통해 해결한 측면이 많았다.

인터넷이 대중화된 이후 뉴스 유통 알고리즘을 이윤을 추구하는 기업이 개발과 업그레이드 주도권을 장악해 왔다. 이런 상황에서 최초로 공공영역에서 비상업적 목표 아래 기사 배열 알고리즘 개발을 시도한 것은 큰 의미를 지닌다고 할 수 있다. 또한, 단순 개발에 그치지 않고 관련한 개발 내용을 모두 공개하여

검증받고 개선 방향을 함께 논의하려 한 측면도 평가받을 수 있다. 기술의 개발 과정에 저널리즘 전공 학자, 컴퓨터 사이언스 학자, 언론사 디지털 부문 경력을 쌓은 기자, 알고리즘 개발 IT 기업 등 다양한 분야의 전문가들이 참여하여 서로 의견을 주고받으면 보다 나은 방안에 대해 논의한 점도 의미를 부여할 수 있다.

포털 등 플랫폼 기업이 기사 배열을 알고리즘에 의해 처리한다고 하지만, 알고리즘에 의한 기사 배열이 어떻게 이뤄지는지에 대한 정보는 거의 주어지지 않아서 이에 대한 일반의 이해가 낮은 것이 사실이다. 포털 등 플랫폼 외부에서는 이에 대한 접근과 평가가 어려웠다. 뉴스 트러스트 프로젝트는 알고리즘에 의한 배열이 실제로 어떻게 설계자와 운영자들의 선호와 가치평가를 반영하는 구조인지를 밝혀내고, 그러한 구조를 언론계와 일반에게 공개하는 프로젝트로서의 의미를 지닌다고 할 수 있다.

하지만, 여러 의미에도 불구하고 뉴스 트러스트 알고리즘은 미완성이다. 앞서 지적한 한계와 동시에 기사의 자동 배열 작업은 특정 기간 집중적으로 진행한다고 완결되는 것이 아니기 때문이다. 지속적으로 개발을 진행하면서 겪게 되는 문제들을 외부에 알리고 이의 해결을 위해 함께 협력하는 것이 앞으로 더욱 중요하다고 할 수 있다. 신뢰라는 가치는 공학적으로 규정하기 굉장히 어렵다. 사람들이 각기 다른 기준을 갖고 있기 때문이다. 신뢰도를 측정하여 기사를 자동 배열한다는 것은 다양한 가치와 관점들을 어떻게 조율해 나가는지가 핵심이다.

04

뉴스 기사 신뢰 요인에 관한 이용자 인식조사

뉴스 트러스트 위원회의 알고리즘 개발 사례 검토를 통해 신뢰도를 기계적으로 측정하는 작업의 어려움과 한계를 대략적으로 알 수 있었다. 이에 더해 실제 이용자들이 뉴스 기사를 읽을 때 어떠한 부분을 통해 신뢰 여부를 결정하는지 그 요인을 알아보고자 했다. 이를 통해 향후 뉴스 기사 신뢰도 측정 방안을 도출하는 데 참고가 될 수 있도록 했다.

1. 조사 설계

1) 조사 목적

본 연구는 온라인 뉴스로 인해 저널리즘 영역에 발생한 다양한 문제점 중에서 뉴스 신뢰에 관한 주제를 다룬다. 월드와이드웹(WWW)이 처음 만들어진 것은 1989년이지만 국내에서 웹 기반 온라인 뉴스 서비스가 이루어진 것은 1990년도 중반부터이다. 국내에서는 중앙일보가 1995년 3월 인터넷 뉴스를 시작하였고, 이후 다양한 전통 언론사가 인터넷 분야를 자회사로 독립시키며 온라인 뉴스 유통에 참여하기 시작하였다. 온라인은 언론사의 다양한 영역에서 변화를 불러왔다. 웹은 대중에게 자율적인 커뮤니케이션 도구를 주었고, 소수 엘리트 취재원에 집중하는 것에서 벗어날 가능성을 제시했다(De Keyser & Raeymaeckers,

2012). 하이퍼링크는 뉴스를 읽는 방식의 변화를 통해 종이신문과는 본질에서 다른 방식의 온라인 뉴스를 만들었다(De Maeyer, 2012). 소셜미디어는 사실이 전달되는 방식을 변화시켰으며, 정치와 언론의 관계를 바꾸어놓기도 했다(Hermida, 2012; Verweij, 2012). 하지만 가장 큰 변화는 디지털 광고단가 하락으로 인한 뉴스 유통방식의 변화였다.

초기 온라인 뉴스는 상대적으로 안정적인 광고 수익을 올렸다. 이후 광고단가는 지속해서 하락하기 시작한다. 프로그램 광고는 도달하고자 하는 소비자를 정확하게 타겟팅하기 시작했으며, 알고리즘에 기반하여 대량으로 이루어지는 광고 집행은 광고단가의 하락을 가져왔다. 광고단가의 하락은 뉴스 기업에게는 재앙에 가까운 결과를 가져왔다. 점차 떨어지는 광고비용을 만회하기 위해 웹 페이지에 다양한 형태의 광고를 삽입하기 시작했다. 하지만 이는 독자들에게도 재앙으로 다가왔다. 수익을 내려는 뉴스 기업들의 노력은 쓸모없고 짜증나는 배너광고, 팝업, 팝언더, 페이지 테이크오버, 끌 수 없는 자동재생 영상, 트래킹 도구, 쿠키 등을 포함한 다른 기괴한 것들로 채워진 웹이라는 결과로 나타났다. 이미 떨어진 광고 수익을 만회할 방법이 없는 뉴스 조직들은 광고 노출보다는 특정 고객을 대상으로 하는 스폰서 광고나 캠페인 기사 작성에 치중하게 되었다. 온라인 저널리즘은 독자들의 요구를 충족시키지 못했고 신뢰의 문제를 가져왔다.

온라인 뉴스 생산·유통·소비 과정에서 사용하는 새로운 기술은 신뢰의 문제를 더욱 악화시키고 있다. 그 중심에는 미디어 영역에서 활용하고 있는 알고리즘이 있다. 알고리즘은 특정한 결과를 산출해내는 규칙이나 명령의 집합으로 뉴스의 생산뿐만 아니라 유통소비 영역에도 영향을 미치고 있다. 딥페이크(deepfake)는 알고리즘을 통해 영상 정보를 조작함으로써 허위정보를 통한 신뢰의 문제를 가져올 소지가 있으며, GPT는 알고리즘에 기반한 자동화된 언어생성 모델로 무분별한 허위정보 생성에 악용될 가능성이 있다. 두 가지 문제 모두 온라인 뉴스의 신뢰와 관련되어 있지만, 온라인 뉴스에서 알고리즘이 영향을 미치는 주요 영역은 유통소비 부분이다. 국내 온라인 뉴스 소비의 대부분은 포털에

의존하고 있으며, 포털은 어떤 뉴스가 중요하고 어떤 뉴스가 사람들에게 노출되어야 하는지 결정한다. 이는 과거 언론사에서 사람이 하던 영역이지만 포털은 사람이 가지는 편향을 제거한다는 명목으로 알고리즘에 뉴스 기사 배열을 의존하기 시작했다.

기존 일반적으로 사용하는 추천 알고리즘은 이용자 행동에 기반하여 개별 항목을 추천한다. 하지만 단순 상품이 아닌 뉴스에 이용자 기반 알고리즘을 적용하는 것이 가져오는 문제점은 없는가? 만약 이용자 행동 기반의 알고리즘이 이용자의 선호를 강화하는 방향으로 작동한다면 사회적으로 중요한 의미가 있는 뉴스보다 이용자가 선호하는 자극적인 뉴스가 알고리즘상 더 높은 평가를 받을 수도 있지 않을까? 하지만 알고리즘에 관한 공개적 논의는 여전히 제한적일 뿐만 아니라 뉴스가 가지는 가치를 계량화하여 측정하고 평가하는 알고리즘을 만들기 위한 시도도 시험적인 수준에만 머물렀다. 본 연구는 알고리즘 뉴스 배열에 대한 이용자 인식을 살펴보고, 뉴스 배열 알고리즘에 적용할 수 있는 신뢰요인이 무엇인지를 확인하고자 했다. 이를 위해 개별 기사와 뉴스 배열에서 신뢰에 영향을 미치는 요인을 확인하는 실험을 했다. 이후 알고리즘을 포함한 뉴스 리터러시 교육의 필요성을 살펴보았다. 최종적으로 뉴스 알고리즘과 저널리즘 신뢰도에 관한 논의가 나아갈 방향에 대해서 논의하고자 한다.

2) 표본 및 설문조사

뉴스 기사 신뢰 요인에 관한 이용자 인식조사는 전문 조사기관인 마켓링크의 온라인 조사를 통해 이루어졌다. 표본은 성별·연령별·지역별 인구비례 할당 방식을 사용했다. 설문조사는 2020년 9월 15일부터 2020년 9월 21일까지 7일간 실시하였다. 설문 발송은 20,512명에게 이루어졌고, 그중에서 5.8%인 1,199명이 설문에 참여하였다. 참여한 1,199명 응답자 중에서 표본 할당에 맞지 않아 제외한 응답자가 66명이었고, 조사 대상에 해당하지 않아 참여하지 못한 응답자

가 83명이었다. 설문을 완료하지 못하고 중도 포기한 응답자는 50명이었다. 최종 설문조사를 완료한 응답자는 1,000명으로 전체 발송 기준 4.9%였다. 전체 응답자 특성은 [표 4-1]과 같다.

[표 4-1] 설문조사 참여자

(단위: 명, %, n=1,000)

| | | 사례 수 | 비율 |
|----|--------------------|-------|-------|
| 성 | 남성 | 512 | 51.2 |
| | 여성 | 488 | 48.8 |
| 연령 | 만 20~29세 | 224 | 22.4 |
| | 만 30~39세 | 224 | 22.4 |
| | 만 40~49세 | 272 | 27.2 |
| | 만 50~59세 | 280 | 28.0 |
| 학력 | 고졸 이하 | 163 | 16.3 |
| | 대졸 이하 | 718 | 71.8 |
| | 대학원 재학 이상 | 119 | 11.9 |
| 소득 | 300만 원 미만 | 247 | 24.7 |
| | 300만 원 ~ 500만 원 미만 | 330 | 33.0 |
| | 500만 원 이상 | 423 | 42.3 |
| 합계 | | 1,000 | 100.0 |

2. 조사 결과

1) 포털 뉴스 이용 일반

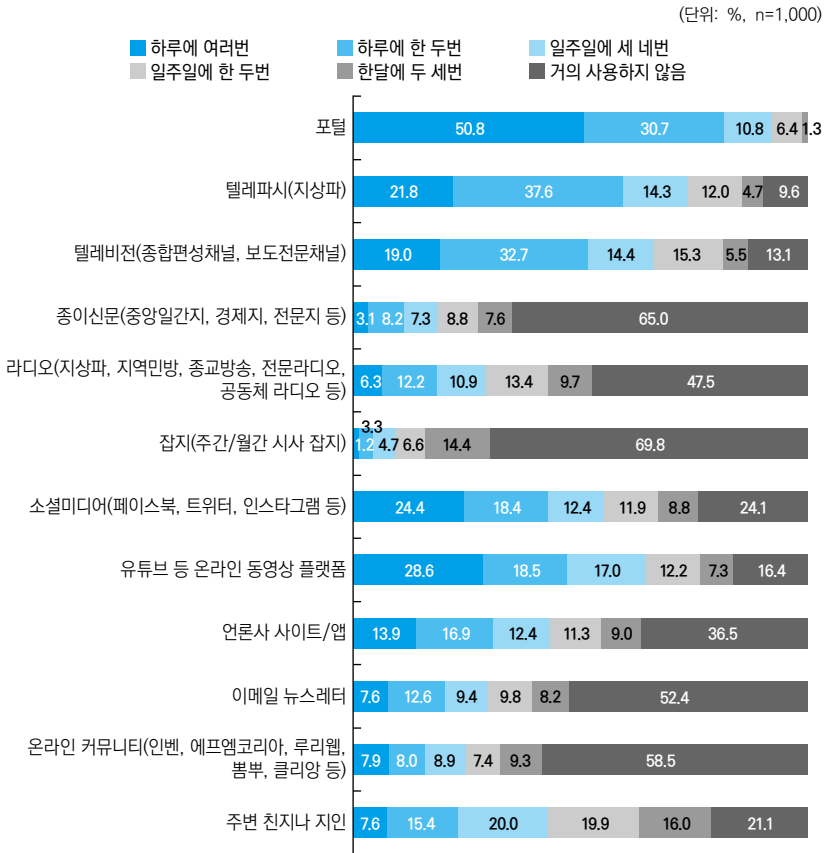
- 뉴스를 접하는 주요 이용경로는 ‘포털’, ‘온라인 동영상 플랫폼’, ‘소셜미디어’ 순으로 나타났으며, 해당 경로는 하루에 여러 번 이용한다는 응답도 여러 경로 중 가장 높은 수준
- ‘잡지’, ‘종이신문’, ‘라디오’ 같은 전통 미디어는 뉴스를 접하는 주요 경로 중 적극적 이용자가 낮은 수준일 뿐만 아니라, 비이용자도 절반 이상으로 높게 나타남
- 포털을 통한 뉴스를 매일 이용한다는 응답은 남성 82.6%, 여성 80.3%였으며, 50대가 가장 높은 87.5%로 연령대가 증가할수록 높은 이용률을 보임
- 포털 뉴스 이용 목적은 ‘정보를 얻기 위함’이 가장 높았으며, ‘개인적 즐거움’이나 ‘지루함을 해소’, ‘타인과 정보를 공유하고 대화하기 위한 목적’ 순으로 나타남
- 포털 이용 동기로 ‘최신 뉴스를 빠르게 알 수 있어서’, ‘실시간으로 뉴스를 접할 수 있어서’, ‘속보를 접할 수 있어서’, ‘모바일에서 확인하기 편리해서’, ‘모든 뉴스를 한 곳에서 접할 수 있어서’와 같은 응답이 있었으며, 포털 신뢰도와 관련된 응답이 가장 낮은 순위를 차지함
- 포털 이용자의 75.8%가 모바일 기기를 통해 이용한다고 답했으며, PC를 이용한다는 응답은 24.2%에 불과했음

(1) 주요 뉴스 이용경로별 이용자 참여 수준

기존 조사 결과는 국내 뉴스 이용경로 대부분을 포털이 차지한다고 이야기하고 있다. 본 연구에서 시행한 설문조사 결과도 포털이 뉴스 이용에서 높은 이용자 참여 수준을 가진다는 사실을 확인해준다. 뉴스를 접하기 위한 다양한 경로를 어느 정도 이용하는지 묻은 결과 포털을 매일 이용하는 응답자는 81.5%(하루에

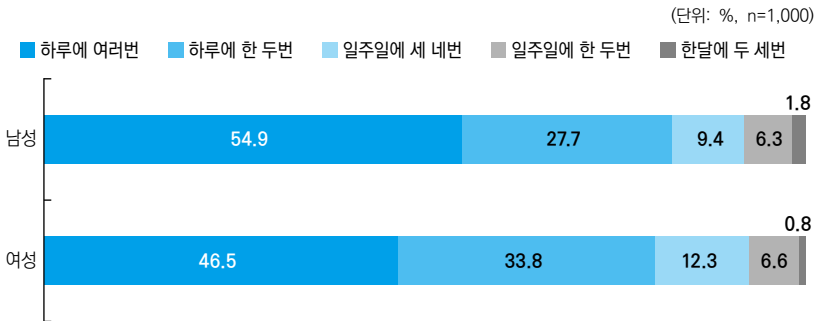
여러 번 50.8% + 하루에 한두 번 30.7%)로 전체 뉴스 이용경로에서 가장 높은 수준을 보였다. 지상파 방송을 매일 이용하는 응답자는 59.4%(하루에 여러 번 21.8% + 하루에 한두 번 37.6%), 종합편성 및 보도전문채널은 51.7%(하루에 여러 번 19.0% + 하루에 한두 번 32.7%)로 뒤를 이었다. 포털과 텔레비전에 이은 주요 뉴스 이용경로가 유튜브와 소셜미디어라는 사실도 주목할 만했다. 유튜브를 매일 이용한다는 응답은 47.1%(하루에 여러 번 24.4% + 하루에 한두 번 18.5%), 소셜미디어는 42.8%(하루에 여러 번 24.4% + 하루에 한두 번 18.4%)였다.

〈그림 4-1〉 경로별 뉴스 이용 빈도



반면, 종이신문이나 잡지와 같은 출판 미디어를 통한 이용 수준은 매우 낮은 수준으로 다양한 뉴스 이용경로 중 최하위권을 차지했다. 가장 낮은 수준을 기록한 잡지는 거의 이용하지 않는다는 응답자가 69.8%에 달했다. 종이신문은 거의 사용하지 않는다는 응답자가 65.0%였으며, 매일 이용한다는 응답도 11.3%(하루에 여러 번 3.1% + 하루에 한두 번 8.2%)에 불과한 것을 확인할 수 있었다. 라디오를 거의 사용하지 않는다는 응답자는 47.5%였으며, 매일 이용한다는 응답은 18.5%(하루에 여러 번 6.3% + 하루에 한두 번 12.2%)에 불과했다. 이러한 결과는 텔레비전을 제외한 전통적인 4대 미디어의 영향력이 매우 미미한 수준까지 떨어졌다는 사실을 보여준다. 동시에, 뉴스 이용의 주요 경로는 여전히 포털이지만 온라인 동영상 플랫폼과 소셜미디어가 새로운 뉴스 이용의 주요 경로로 부상하고 있음을 알 수 있다. 이용자 참여 수준을 예측해볼 수 있는 하루에 여러 번 이용하는 응답자를 살펴보면 포털(50.8%), 온라인 동영상 플랫폼(28.6%), 소셜미디어(24.4%)가 가장 높은 순위를 기록하고 있다.

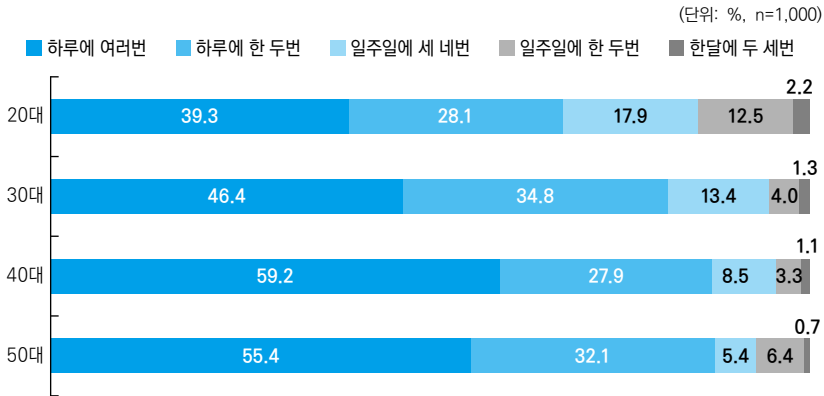
〈그림 4-2〉 포털 뉴스 이용 빈도(성별)



포털 뉴스 이용 수준을 성별·연령별로 살펴보았다. 매일 이용하는 응답자를 기준으로 살펴보았을 때, 남성은 82.6%(하루에 여러 번 54.9% + 하루에 한두 번 27.7%), 여성은 80.3%(하루에 여러 번 46.5%, 하루에 한두 번 33.8%)였다. 연령별로 살펴보았을 때 20대의 포털 뉴스 이용 정도는 다른 세대와 달리 확연히 낮았다. 매일 이용하는 응답자는 67.4%(하루에 여러 번 39.3% + 하루에 한두

번 28.1%)에 불과했다. 매일 이용하는 응답자는 연령대가 높아질수록 증가하는 경향을 보였다. 30대는 81.3%(하루에 여러 번 46.4% + 하루에 한두 번 34.8%), 40대는 87.1%(하루에 여러 번 59.2% + 하루에 한두 번 27.9%), 50대는 87.5%(하루에 여러 번 55.4% + 하루에 한두 번 32.1%)였다.

〈그림 4-3〉 포털 뉴스 이용 빈도(연령별)



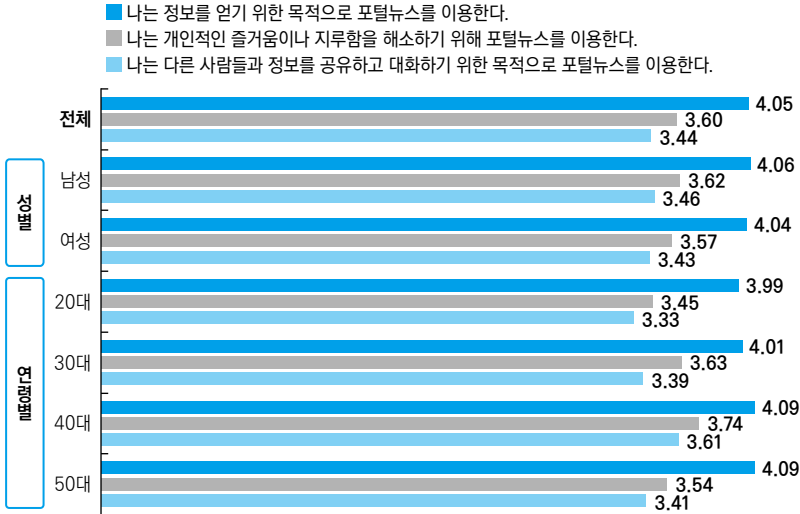
(2) 포털 이용 목적 및 동기

뉴스를 이용하는 목적은 다양하다. 정보를 얻기 위한 주요 수단이기도 하지만 지루한 시간을 해소하거나 다른 사람과 사교를 위한 정보 획득 및 공유 목적으로 이용하기도 한다. 이와 같은 뉴스 이용 목적 중 자신의 포털 뉴스 이용 목적은 어디에 해당하는지 물었다. ‘정보를 얻기 위한 목적으로 포털 뉴스를 이용한다.’는 응답은 5점 만점 기준에 4.05점을 받았다. ‘개인적인 즐거움이나 지루함을 해소하기 위해 포털 뉴스를 이용한다.’는 응답은 3.60점을 받았다. ‘다른 사람과 정보를 공유하고 대화하기 위한 목적으로 포털 뉴스를 이용한다.’는 응답은 3.44점을 받았다. 뉴스 이용 목적은 성별·연령별 구분에서도 같은 순서를 보였다. 성별 차이는 크지 않은 편이었고, 연령별 차이는 다른 연령대보다 40대에서 약간 높은 수치를 보였다. 40대는 정보 목적 이용이 4.09점, 즐거움이나 지루함 해소가 3.74점, 타인과 정보공유나 대화 목적이 3.61점으로 나타났다. 그럼에도

불구하고 다른 연령대와 차이는 통계적으로 유의하지 않았다($F_{성별} = 0.17, p = n.s.$; $F_{연령별} = 1.42, p = n.s.$).

〈그림 4-4〉 포털 뉴스 이용 목적

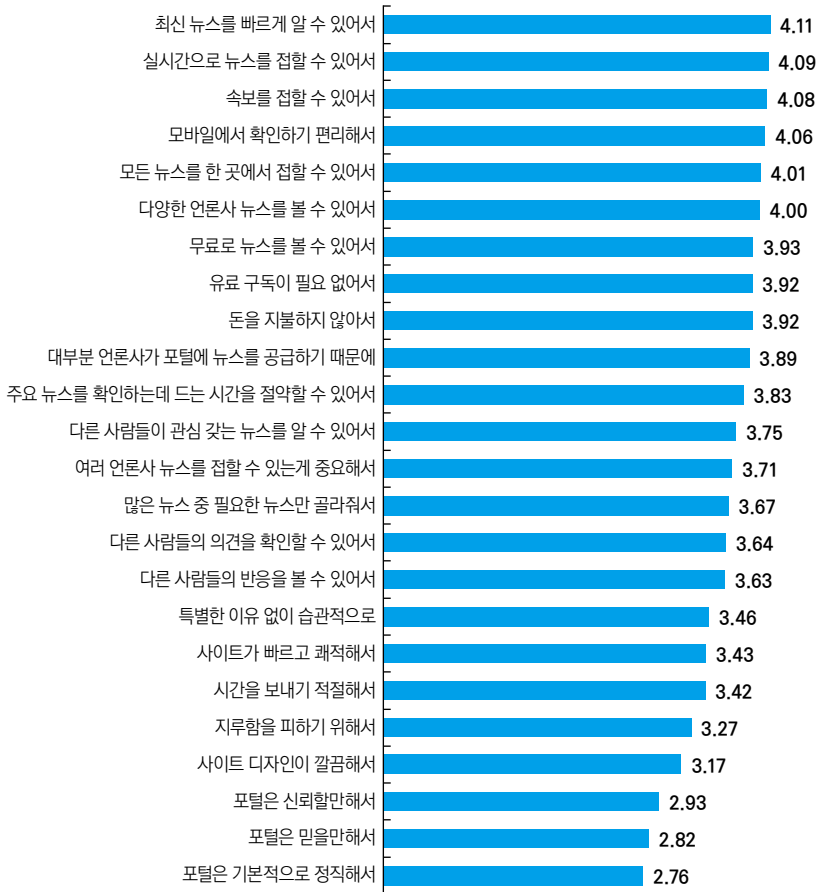
(단위: 점, n=1,000)



다양한 포털 뉴스 이용 동기 질문에서 가장 높은 응답을 받은 요인은 빠른 뉴스의 이용과 관련 있는 요인이었다. ‘최신 뉴스를 빠르게 알 수 있어서’, ‘실시간으로 뉴스를 접할 수 있어서’, ‘속보를 접할 수 있어서’ 세 문항이 5점 만점 기준 각각 4.11점, 4.09점, 4.08점으로 높은 순위를 차지했다. 다음으로 높은 응답을 받은 질문은 ‘모바일에서 확인하기 편리해서(4.06점)’, ‘모든 뉴스를 한 곳에서 접할 수 있어서(4.01점)’, ‘다양한 언론사 뉴스를 볼 수 있어서(4.00점)’였다. 반면 가장 낮은 응답을 받은 질문은 신뢰와 관련된 문항이었다. ‘포털은 신뢰할 만해서’, ‘포털은 믿을 만해서’, ‘포털은 기본적으로 정직해서’라는 질문은 각각 2.93점, 2.82점, 2.76점을 받았다. 이러한 결과는 포털을 이용하는 목적은 뉴스를 통한 정보 획득에 있으나, 포털을 신뢰하기 때문이라기보다는 정보를 빠르고 편리하게 획득할 수 있기 때문이라는 사실을 보여준다.

〈그림 4-5〉 포털 뉴스 이용 동기

(단위: 점, n=1,000)

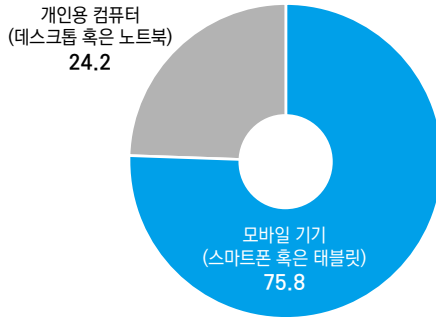


(3) 포털 뉴스 이용 방식

포털을 통해 뉴스를 이용할 때 주로 어떤 기기를 사용하는지 물었다. 스마트폰이나 태블릿 같은 모바일 기기를 통해 이용한다는 응답이 75.8%, 데스크톱이나 노트북과 같은 개인용 컴퓨터를 통해 이용한다는 응답이 24.2%였다.

〈그림 4-6〉 포털 뉴스 이용 시 주로 사용하는 기기

(단위: %, n=1,000)



단순히 모바일을 통한 이용이 높다는 조사 결과 자체가 중요하다기보다는 모바일을 통해 포털 뉴스를 이용하는 경우 알고리즘을 통한 뉴스 배열에 영향을 받을 가능성이 크다는 사실이 더 중요하다. 모바일 기기는 개인용 컴퓨터보다 화면이 작기에 직접적인 선택보다 플랫폼이 제공하는 콘텐츠 큐레이션의 역할이 뉴스 이용에서 중요한 비중을 차지하기 때문이다.

2) 알고리즘 뉴스 배열에 관한 인식

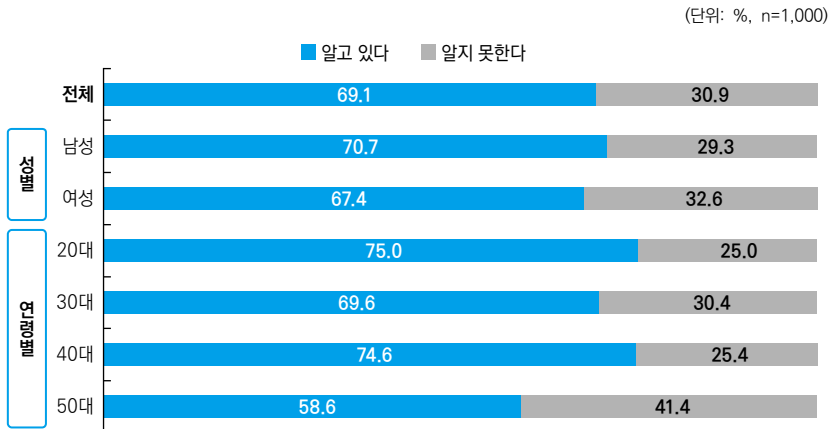
- 전체 응답자 10명 중 7명은 포털이 뉴스 배열에 알고리즘을 활용한다는 사실을 인지하고 있었음
- 알고리즘이 상업성과 편향성을 가진다는 질문에 동의하는 응답자가 많았고, 알고리즘의 신뢰성에는 상대적으로 낮은 평가를 보임
- 알고리즘 뉴스 배열이 다양한 뉴스를 제공하고 중요한 뉴스를 볼 수 있게 한다는 긍정적인 평가와 함께 알고리즘으로 인해 다른 방식의 뉴스 이용이 증가했으며, 내 관심에서 벗어난 주제를 접하기 어려워졌다는 의견이 공존했음. 반면, 일반적으로 논의되어 온 개인정보 유출이라든지 원하는 정보를 접하기 어려워질 것이라는 두려움은 낮은 수준임

- 신뢰할 만한 뉴스 배열 주체는 기계라는 응답이 65.7%로 사람보다 우세했으며, 심층성을 제외한 공정성, 투명성, 다양성, 객관성, 정확성, 화제성에서 모두 기계가 업무를 더 잘 수행할 것이라 봄

(1) 포털의 알고리즘 뉴스 배열 인지 여부

포털은 이용자에게 어떤 기사를 보여줄지를 알고리즘에 의존한다. 종이신문 시절 어떤 기사가 신문의 1면에 나와야 하는지를 결정하는 것은 인간이었다. 충분한 경력을 쌓은 기자가 현장에서 보내는 기사를 취합하고 보도 방향을 결정했다. 하지만 포털은 인간의 손을 거쳐 이루어지던 뉴스 배열을 알고리즘으로 처리한다고 밝히고 있다. ‘다음’은 오래전부터 루빅스(RUBICS)라고 부르는 알고리즘에 기반하여 개인에게 가장 적합한 맞춤형 기사를 제공했다. ‘네이버’도 2019년부터 에어스(AiRS)라고 부르는 알고리즘 기반 자동 추천 기사를 전면 적용했다. 네이버가 2017년 K리그에 부정적인 기사를 보이지 않게 해달라는 청탁을 받고 기사 배열을 변경한 임원에게 징계 조치를 한 이후 이루어진 일이다.

〈그림 4-7〉 알고리즘 뉴스 배열 인지 여부



과연 응답자들은 알고리즘에 대해 어느 정도 인지하고 있을까? 질문 전 응답자에게 “현재, 네이버, 다음, 구글, 유튜브, 페이스북과 같은 포털과 소셜미디어는 사람이 직접 배열하던 뉴스를 기계적 알고리즘을 통해 개인 맞춤형으로 자동 배열하고 있습니다.”라는 내용을 포털 뉴스 배열 화면과 함께 제시했다. 이후 “뉴스를 접하기 위해 사용하는 다양한 서비스에서 알고리즘을 사용한다는 사실을 인지하고 있으십니까?”라고 물었다. 전체 응답자 기준으로 보았을 때 알고리즘 사용을 인지하고 있다는 응답자는 69.1%로 10명 중 7명 정도였다. 알고리즘에 대한 이용자의 인지 수준이 상당히 높은 수준임을 확인할 수 있었다.

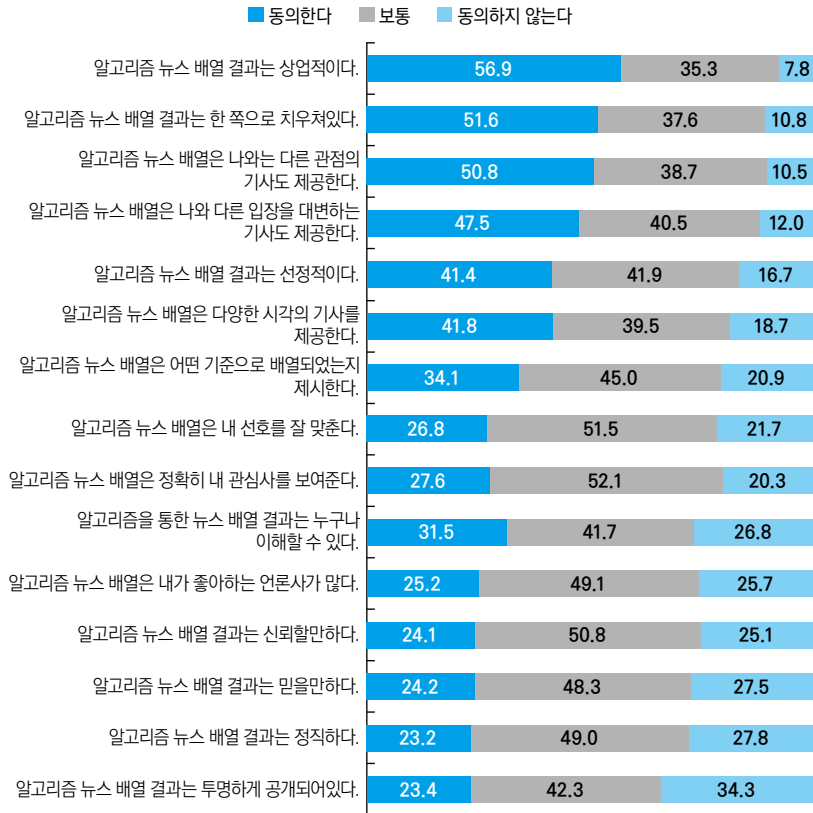
성별에 따른 알고리즘 뉴스 배열 인지 여부를 살펴보면 남성은 70.7%, 여성은 67.4%가 알고 있다고 답했다. 하지만 성별에 따른 차이는 유의하지 않았다($\chi^2(1, N = 1,000) = 1.26, p = n.s.$). 연령에 따른 인지 여부는 통계적으로도 유의한 차이를 보였다($\chi^2(3, N = 1,000) = 22.12, p < 0.001$). 알고리즘 뉴스 배열 인지가 가장 높은 연령대는 20대로 75.0%가 알고 있다고 답했다. 두 번째로 높은 연령대는 40대로 74.6%였다. 오히려 30대가 40대보다 알고리즘 뉴스 배열에 대한 인지 정도가 낮았다. 가장 낮은 것은 50대로 58.6%는 알고 있다고 답했으나 41.4%는 알지 못한다고 응답했다. 즉, 평균적으로 10명 중 7명이 알고리즘을 활용하여 뉴스를 배열한다는 사실을 알고 있었으나 50대에서는 알고 있다는 응답이 다른 연령대에 비해 낮은 수준이었다.

(2) 알고리즘 뉴스 배열에 관한 인식과 영향력

포털 뉴스 이용자가 알고리즘 뉴스 배열에 관해 어떤 인식을 하고 있는지 살펴보았다. 알고리즘 뉴스 배열에 관한 다양한 질문에 어느 정도 동의하는지를 5점 기준으로 물었다.

〈그림 4-8〉 알고리즘 뉴스 배열에 관한 인식

(단위: %, 점, n=1,000)

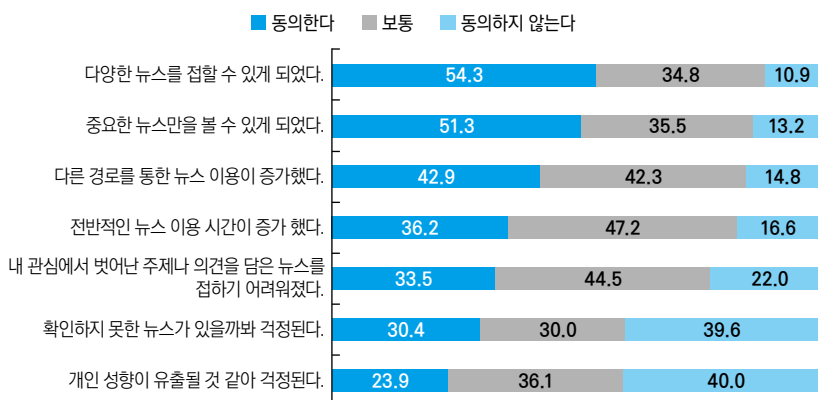


가장 많은 동의를 받은 응답은 알고리즘의 상업성에 관한 내용이었다. ‘알고리즘 뉴스 배열 결과는 상업적이다.’라는 질문에 56.9%, ‘알고리즘 뉴스 배열 결과는 한쪽으로 치우쳐 있다.’라는 질문은 51.6%, ‘알고리즘 뉴스 배열은 나와 다른 관점의 기사도 제공한다.’라는 질문에는 50.8%의 응답자가 동의했다. ‘알고리즘 뉴스 배열은 나와 다른 입장을 대변하는 기사도 제공한다(47.5%).’, ‘알고리즘 뉴스 배열 결과는 선정적이다(41.4%).’, ‘알고리즘 뉴스 배열은 다양한 시각의 기사를 제공한다(41.8%).’와 같은 응답이 뒤를 이었다. 반면 낮은 동의를 받은 응답은 결과의 투명성이나 신뢰와 관련된 내용이었다. ‘알고리즘 뉴스 배열 결과

는 투명하게 공개되어 있다.’라는 질문에 동의하지 않는 응답자는 34.3%로 동의한다는 응답 23.4%보다 높았다. ‘알고리즘 뉴스 배열 결과는 정직하다.’, ‘알고리즘 뉴스 배열 결과는 믿을 만하다.’, ‘알고리즘 뉴스 배열 결과는 신뢰할 만하다.’라는 질문 모두 동의하지 않는다는 응답자가 동의한다는 응답자에 비해 많았다. 전반적으로 알고리즘이 편향성을 가진다는 데 동의하는 응답자가 많았고, 알고리즘의 신뢰성을 부정적으로 바라보는 응답자가 많음을 알 수 있었다.

〈그림 4-9〉 알고리즘 뉴스 배열의 영향력

(단위: %, n=1,000)



알고리즘 뉴스 배열이 뉴스 이용과 관련해 어떤 영향을 미쳤는지를 물었다. 가장 많은 동의를 얻은 문항은 ‘다양한 뉴스를 접할 수 있게 되었다.’로 54.3%, ‘중요한 뉴스만을 볼 수 있게 되었다.’는 51.3%, ‘다른 경로를 통한 뉴스 이용이 증가했다.’는 42.9%, ‘전반적인 뉴스 이용 시간이 증가했다.’는 36.2%, ‘내 관심에서 벗어난 주제나 의견을 담은 뉴스를 접하기 어려워졌다.’는 33.5%의 응답자가 동의했다. 반면 ‘개인성향이 유출될 것 같아 걱정된다.’는 동의하지 않는 응답자가 40.0%였으며, ‘확인하지 못한 뉴스가 있을까봐 걱정된다.’는 응답은 39.6%가 동의하지 않는다고 답했다. 알고리즘 뉴스 배열의 영향력에 대한 질문은 전반적으로 긍정적인 평가가 더 우세했다. 다양한 뉴스와 중요한 뉴스만을 접할 수

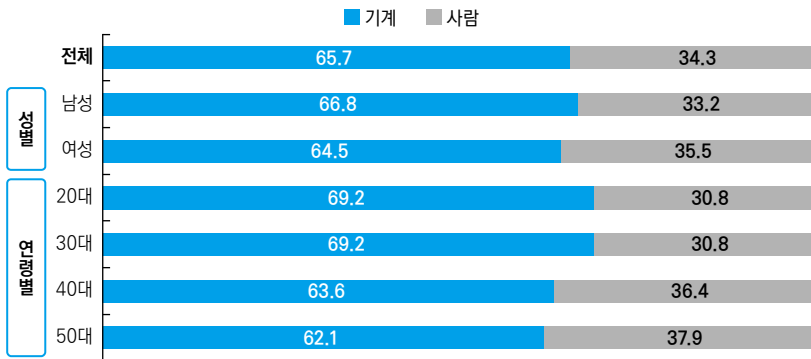
있고 전반적 뉴스 이용시간이 증가했다는 답변에 동의하는 응답이 많았다. 하지만 알고리즘으로 인해 다른 방식의 뉴스 이용이 증가했으며, 내 관심에서 벗어난 주제를 접하기 어려워졌다는 부정적인 응답도 일정 비율의 응답자가 동의했다. 일반적으로 알려진 알고리즘으로 인한 개인정보 유출, 원하는 정보를 접하지 못할 것이라는 두려움은 낮은 수준인 것을 확인할 수 있었다.

(3) 뉴스 배열에서 기계와 사람에 관한 평가

실제 뉴스 배열을 담당하는 주체가 기계일 때와 사람일 때를 비교한다면 어느 쪽이 더 신뢰할 만하다고 생각하는지 물었다. 기계가 더 신뢰할 만하다는 응답은 65.7%로 사람이 더 신뢰할 만하다는 응답 34.3%에 비해 높았다. 즉, 3명 중 2명은 기계의 뉴스 배열이 사람의 배열보다 믿을 만하다는 반응을 보였다.

〈그림 4-10〉 신뢰할 만한 뉴스 배열 주체

(단위: %, n=1,000)

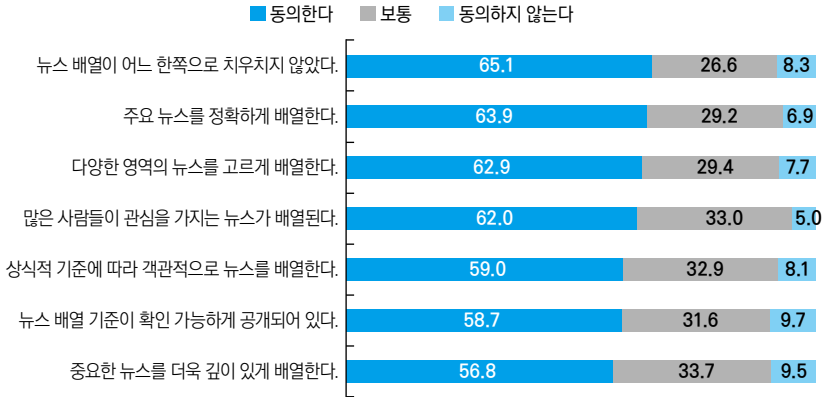


성별 차이는 기계가 신뢰할 만하다고 답변한 남성이 66.8%, 여성이 64.5%로 나타나 남성이 기계를 신뢰한다는 답변이 약간 높았으나 통계적으로 유의한 차이는 아니었다($\chi^2(1, N = 1,000) = 0.56, p = n.s.$). 연령별로는 20대와 30대에서 기계가 더 신뢰할 만하다는 응답이 69.2%로 다른 연령대에 비해 높았으며 50대는 기계가 더 믿을 만하다는 응답이 62.1%로 다른 연령대에 비해 낮았다.

하지만 연령별 차이는 통계적으로 유의하지는 않았다($\chi^2(3, N = 1,000) = 4.53, p = n.s.$).

〈그림 4-11〉 뉴스 배열에서 중요하다고 생각하는 기준

(단위: %, n=1,000)

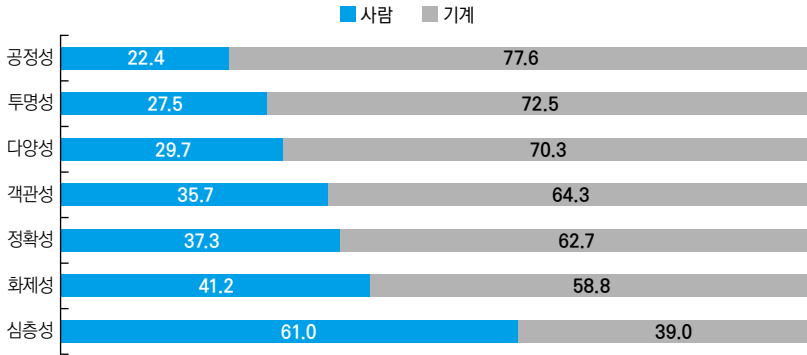


공정성, 다양성, 정확성, 심층성, 투명성, 객관성, 화제성에 관한 기준이 뉴스 배열에서 얼마나 중요하다고 생각하는지 물었다. 가장 다수가 동의한 응답은 ‘뉴스 배열이 어느 한쪽으로 치우치지 않았다.’라는 공정성 기준으로 전체 응답자의 65.1%가 동의했다. ‘주요 뉴스를 정확하게 배열한다.’라는 정확성 기준은 63.9%로 두 번째를 차지했다. ‘다양한 영역의 뉴스를 고르게 배열한다.’라는 다양성 기준은 62.9%로 세 번째였다. ‘많은 사람들이 관심을 가지는 뉴스가 배열된다.’라는 화제성 기준은 62.0%, ‘상식적 기준에 따라 객관적으로 뉴스를 배열한다.’라는 객관성 기준은 59.0%, ‘뉴스 배열 기준이 확인 가능하게 공개되어 있다.’라는 투명성 기준은 58.7%, ‘중요한 뉴스를 더욱 깊이 있게 배열한다.’라는 심층성 기준은 가장 낮은 56.8%의 응답자가 동의한다고 답했다. 모든 기준에 절반 이상의 응답자가 중요하다는 데 동의했으나, 가장 많은 응답자가 동의한 기준은 공정성이었으며 가장 적은 응답자가 동의한 기준은 심층성이었다. 즉, 뉴스 배열에서 지켜야하는 가치로 논의되어 왔던 다양한 기준 중에서 이용자가

중요하게 생각하는 것은 ‘공정성’, ‘정확성’, ‘다양성’ 순으로 나타났으며, 부정적인 시각으로 논의되기도 하는 ‘화제성’과 같은 기준도 이용자 입장에서는 높은 중요성을 가지는 배열 기준으로 나타났다.

〈그림 4-12〉 뉴스 배열의 주요 기준을 잘 수행할 수 있는 주체

(단위: %, n=1,000)



뉴스 배열에서 중요하다고 언급되는 기준을 잘 지키며 뉴스를 배열할 수 있는 것은 기계와 사람 중에서 누구라고 생각하는지를 물었다. 응답자들은 심층성을 제외한 모든 기준에서 사람보다 기계가 더 잘할 것으로 생각했다. 심층성에서 기계라는 응답은 39.0%, 사람이라는 응답은 61.0%로 사람이 심층성 있는 기사를 더 잘 배열할 것이라는 판단이 우세했다. 반면 나머지 기준은 모두 기계가 더 잘할 것이라는 답변을 확인할 수 있었다. 각 기준에서 기계라는 응답을 중심으로 살펴보면 공정성 77.6%, 투명성 72.5%, 다양성 70.3%, 객관성 64.3%, 정확성 62.7%, 화제성 58.8%였다.

3) 뉴스 신뢰도와 알고리즘 기사 평가

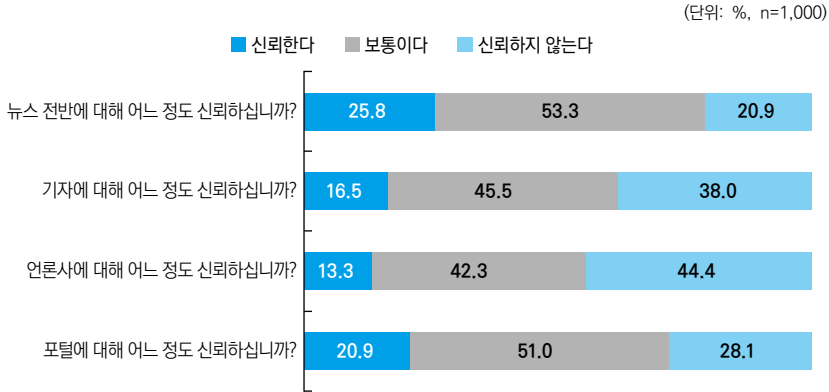
- 뉴스 전반에 대해 신뢰한다는 응답은 25.8%였으나, 기자를 신뢰한다는 응답은 16.5%, 언론사를 신뢰한다는 응답은 13.3%, 포털을 신뢰한다는 응답은 20.9%로 낮게 나타남. 특히, 기자와 언론사는 신뢰하지 않는다는 응답이 각각 38.0%, 44.4%로 높았음
- 신뢰하는 기자나 언론사 유무를 물었을 때, 신뢰하는 기자가 있다는 응답은 12.4%, 신뢰하는 언론사가 있다는 응답은 20.7%
- 신뢰하는 기자 1위는 주진우, 신뢰하는 언론사 1위는 JTBC였으나 신뢰하는 기자나 언론사를 답한 응답자가 극히 소수이기 때문에 신뢰하는 기자나 언론사가 없다는 사실에 주목할 필요가 있음
- 뉴스 신뢰도 평가에 동의한다는 응답자는 86.7%였으며, 그중에서 기계적으로 신뢰도를 평가하는 데 동의한다는 응답자는 73.2%였음
- 기계적 평가에 찬성하는 이유로는 '사람이 가진 편견을 제거할 수 있기 때문'이라는 응답이 가장 많았으며, 반대하는 이유로는 '모든 기사를 같은 기준으로 판단하는 게 어렵기 때문'이라는 응답이 가장 많았음
- 기계적 평가 점수를 포털의 뉴스 배열에 반영하는 것을 찬성하는 응답자는 기계적 평가에 찬성한 응답자 중 95.3%. 찬성 이유로는 '신뢰도를 중심으로 다양한 뉴스를 볼 수 있기 때문'이라는 응답이 가장 많았으며, 반대 이유로는 '신뢰도 평가를 믿을 수 없어서'라는 답변이 가장 많았음

(1) 뉴스 신뢰도

뉴스 이용경로의 포털 집중은 온라인을 통해 뉴스가 유통되기 시작한 이후 지속해서 제기된 문제 중 하나였다. 뉴스 이용이 집중된 상태에서 포털의 뉴스 알고리즘 적용은 뉴스 신뢰와 관련된 문제를 다시 한번 논의하게 하는 계기가 되었다. 로이터의 디지털 뉴스 리포트에서 이루어지는 언론 신뢰도 조사 결과에 따르면, 한국은 뉴스를 신뢰하는지에 관한 질문에 21%가 동의하여 조사 대상

40개 국가 중 최하위를 기록했다. 뉴스 신뢰도와 관련된 문제를 확인하기 위해 현재 응답자가 생각하는 일반적인 뉴스 신뢰도를 살펴보고, 뉴스 이용자가 신뢰를 판단하는 요인이 무엇인지 확인하고자 했다. 또한, 이러한 신뢰 요인을 바탕으로 이루어지는 기계적인 신뢰도 평가를 도입하는 것에 동의하는지를 물었다.

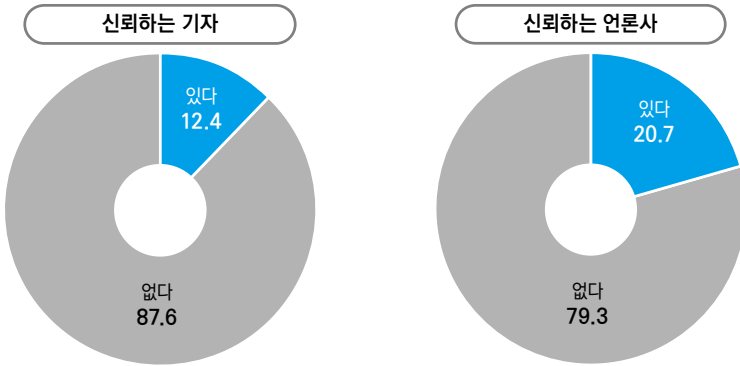
〈그림 4-13〉 뉴스 관련 신뢰도



우선 뉴스에 관련된 다양한 주체의 현재 신뢰도를 확인하고자 했다. 이를 위해 응답자에게 전반적인 뉴스에 대한 신뢰도를 확인한 후, 기자, 언론사, 포털을 어느 정도 신뢰하는지를 물었다. 뉴스 전반에 대해 어느 정도 신뢰하는지 묻는 말에 25.8%가 신뢰한다고 답했으며, 신뢰하지 않는다는 응답은 20.9%였다. 기자, 언론사, 포털에 대한 신뢰도는 뉴스 전반에 대한 신뢰도보다 더 부정적인 결과를 보였다. 기자에 대해 신뢰한다는 응답은 16.5%, 신뢰하지 않는다는 응답은 38.0%였다. 언론사에 대해 신뢰한다는 응답은 셋 중 가장 낮은 13.3%, 신뢰하지 않는다는 응답은 44.4%였다. 포털에 대해 신뢰한다는 응답은 20.9%, 신뢰하지 않는다는 응답은 28.1%였다. 기자와 언론사에 대해서는 신뢰한다는 응답이 낮았지만, 신뢰하지 않는다는 응답이 높다는 점도 두드러지는 결과 중 하나였다.

〈그림 4-14〉 신뢰하는 기자와 언론사 유무

(단위: %, n=1,000)



뉴스와 관련된 다양한 주체가 전반적으로 낮은 신뢰도를 보였다. 그런데도 그중에서 신뢰하는 기자나 언론사가 있는지를 물었다. 신뢰하는 기자가 있다는 응답자는 전체 12.4%였으며, 신뢰하는 언론사가 있다는 응답자는 전체 20.7%였다. 신뢰하는 기자나 언론사가 있다고 응답한 사람들에게 신뢰하는 기자나 언론사 이름이 무엇인지를 개방형 응답으로 요청하였다. 신뢰하는 기자가 있다는 응답은 전체 응답자 1,000명 중 12.4%인 124명이었으며, 그중에서 가장 많은 응답을 받은 기자는 16.1%를 받은 주진우 기자였다(1,000명 중 20명). 박대기 기자는 9.7%(1,000명 중 12명)로 두 번째였으며, 손석희 기자 7.3%(1,000명 중 9명), 서복현 기자 3.2%(1,000명 중 4명)로 나타났다.

신뢰하는 언론사가 있다는 응답은 전체 응답자 1,000명 중 20.7%인 207명이었다. 그 중에서 가장 많은 응답을 받은 JTBC는 20.8%(1,000명 중 43명)로 나타났으며, 조선일보 12.6%(1,000명 중 26명), 한겨레 11.6%(1,000명 중 24명), MBC와 YTN 4.8%(1,000명 중 10명)로 나타났다. 신뢰하는 기자와 언론사에 대한 설문 조사 결과의 해석은 해당 기자와 언론사의 순위가 중요하다기보다는 전체 응답자에서 신뢰하지 않는다고 답한 응답자 숫자가 압도적이고 오직 소수의 응답자만 소수의 기자나 언론사를 인지하고 신뢰한다는 것으로 바라보아야 한다.

[표 4-2] 신뢰하는 기자와 언론사 명단

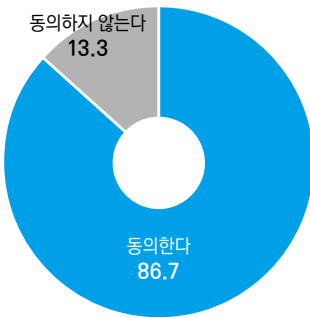
| (단위: 명, %, n=124/207) | | | |
|-----------------------|--------|-------|--------|
| 기자 | 응답자 비율 | 언론사 | 응답자 비율 |
| 주진우 | 16.1 | JTBC | 20.8 |
| 박대기 | 9.7 | 조선일보 | 12.6 |
| 손석희 | 7.3 | 한겨레신문 | 11.6 |
| 서복현 | 3.2 | MBC | 4.8 |
| 김어준 외 4인 | 1.6 | YTN | 4.8 |

(2) 알고리즘 기반 뉴스 신뢰도 평가 여부

뉴스 신뢰도가 전반적으로 낮은 상황에서 하나의 대안으로 뉴스 기사 신뢰도를 평가하는 것에 대한 이용자의 의견을 물었다. 응답자의 86.7%는 뉴스 기사 신뢰도를 평가하는 것에 동의한다고 답했다. 반면 13.3%는 동의하지 않는다고 답해 찬성한다는 응답자가 다수임을 알 수 있었다.

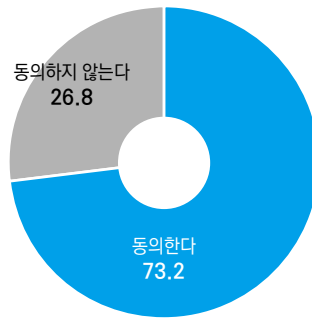
〈그림 4-15〉 뉴스 신뢰도 평가 여부

(단위: %, n=1,000)



〈그림 4-16〉 뉴스 신뢰도의 기계적 평가 여부

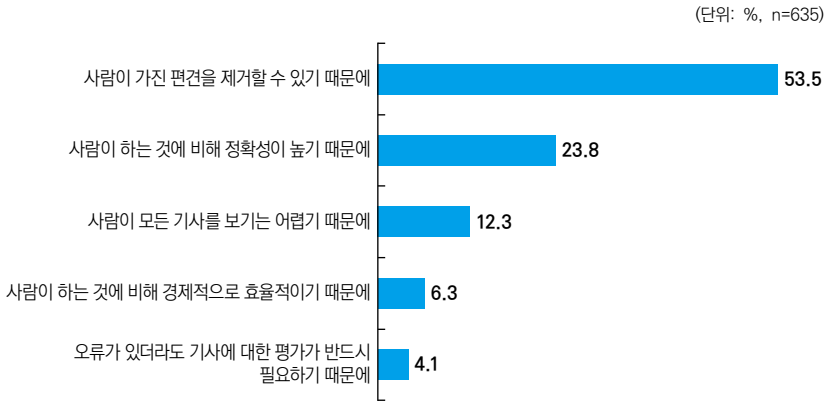
(단위: %, n=867)



찬성한다고 답한 응답자 867명을 대상으로 뉴스 기사 신뢰도를 컴퓨터가 기계적으로 평가하는 것에 동의하는지를 물었다. 컴퓨터가 신뢰도를 평가하는 것에 동의하는 응답자는 73.2%였으며, 동의하지 않는다는 응답자는 26.8%였다. 뉴

스 신뢰도를 기계적으로 평가하는 데 찬성하는 이유 중 가장 많은 선택을 받은 것은 ‘사람이 가진 편견을 제거할 수 있기 때문’이라는 응답으로 53.5%가 동의한다고 답했다. ‘사람이 하는 것에 비해 정확성이 높기 때문’이라는 응답에 23.8%, ‘사람이 모든 기사를 보기는 어렵기 때문’이라는 응답에 12.3%가 동의했다. ‘사람이 하는 것에 비해 경제적으로 효율적이기 때문에’, ‘오류가 있더라도 기사에 대한 평가가 반드시 필요하기 때문’이라는 응답이 각각 6.3%, 4.1%로 나타났다.

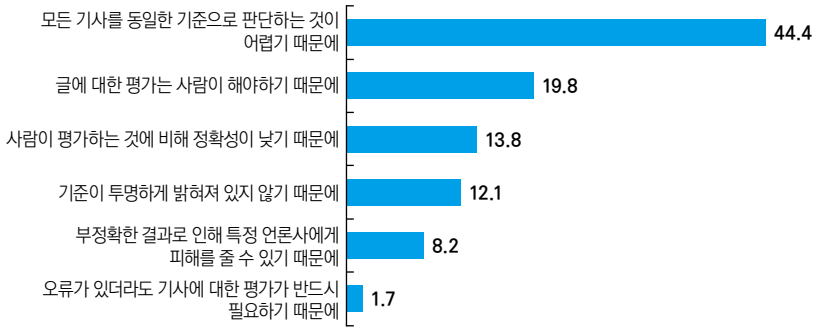
〈그림 4-17〉 뉴스 신뢰도의 기계적 평가에 찬성하는 이유



반대하는 이유로는 ‘모든 기사를 같은 기준으로 판단하는 것이 어렵기 때문’이라는 응답이 44.4%로 가장 많은 동의를 받았다. ‘글에 대한 평가는 사람이 해야 하기 때문’이 19.8%, ‘사람이 평가하는 것에 비해 정확성이 낮기 때문’이 13.8%, ‘기준이 투명하게 밝혀져 있지 않기 때문’이 12.1%, ‘부정확한 결과로 인해 특정 언론사에게 피해를 줄 수 있기 때문’이 8.2%로 나타났다. 기타 의견으로는 ‘기계가 인간의 가치 판단을 따라오지 못하기 때문’이라거나 ‘해킹 우려 때문’이라는 답변이 있었다.

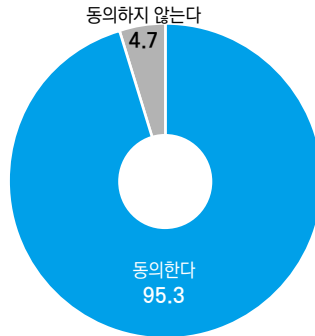
〈그림 4-18〉 뉴스 신뢰도의 기계적 평가에 반대하는 이유

(단위: %, n=232)



〈그림 4-19〉 기계적 평가 점수를 뉴스 배열에 반영할지 여부

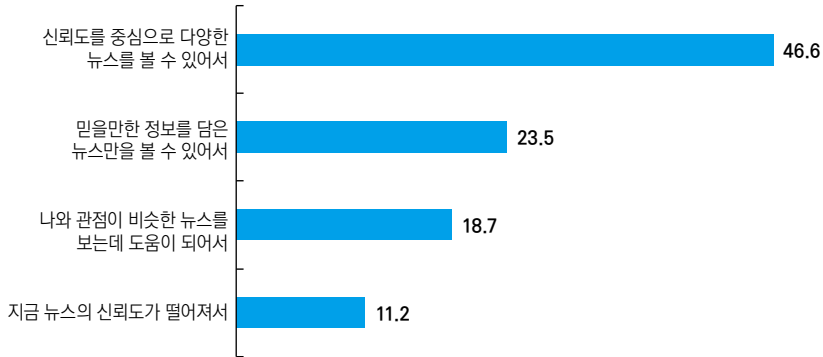
(단위: %, n=635)



뉴스 신뢰도를 기계적으로 평가하는 것에 동의한 635명의 응답자에게 신뢰도 점수를 알고리즘 뉴스 배열 기준에 포함하는 것에 동의하는지를 물었다. 소수의 응답자(4.7%)를 제외한 대부분 응답자(95.3%)가 뉴스 배열 기준에 알고리즘 신뢰도 평가 점수를 포함하는 것에 동의했다. 동의하는 이유를 물었을 때 ‘신뢰도를 중심으로 다양한 뉴스를 볼 수 있어서’라는 응답이 가장 많은 46.6%였다. ‘믿을 만한 정보를 담은 뉴스만 볼 수 있어서’라는 응답이 23.5%, ‘나와 관점이 비슷한 뉴스를 보는 데 도움이 되어서’라는 응답이 18.7%, ‘지금 뉴스의 신뢰도가 떨어져서’라는 응답이 11.2%였다.

〈그림 4-20〉 기계적 평가 점수를 뉴스 배열에 반영하는 것을 찬성하는 이유

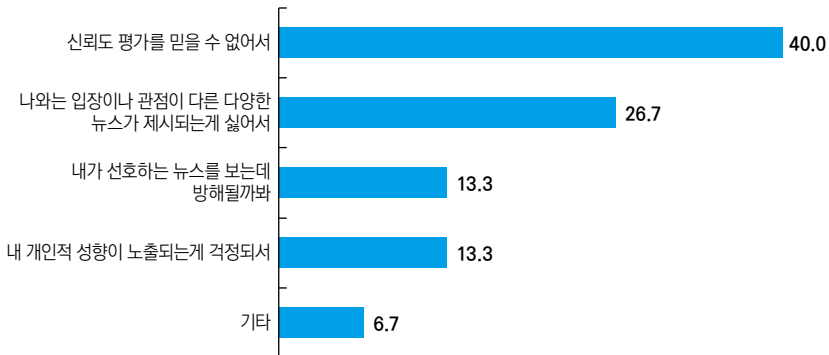
(단위: %, n=605)



반면, 기계적 평가 점수를 뉴스 배열에 반영하는 것에 반대하는 이유로는 ‘신뢰도 평가를 믿을 수 없어서’라는 응답이 40.0%, ‘나와는 입장이나 관점이 다른 다양한 뉴스가 제시되는 게 싫어서’가 26.7%, ‘내가 선호하는 뉴스를 보는데 방해될까 봐’, ‘내 개인적 성향이 노출되는 게 걱정돼서’라는 응답이 13.3%를 차지했다. 기타 응답으로는 ‘신뢰한다고 해서 모든 기사가 다 옳은 것은 아니기 때문에’, ‘다양한 뉴스가 사라질까 봐’와 같은 이유가 있었다.

〈그림 4-21〉 기계적 평가 점수를 뉴스 배열에 반영하는 것을 찬성하는 이유

(단위: %, n=30)



4) 뉴스 신뢰도 평가 요인

- 신뢰도 평가에 영향을 미치는 요인을 ‘기사’, ‘기자’, ‘언론사’, ‘타인’으로 나누어 살펴봤을 때, 가장 중요한 것은 기사 자체 요인
- 출처 확인이나 사실 검증 관련된 기사의 객관성, 중립적인 보도와 관련된 공정성, 협찬 기사의 명확한 표현과 같은 투명성이 뉴스 신뢰도를 판단하는 주요한 기사 요인 특정 주제에 관해 내가 동의하는 보도를 하는지는 신뢰도 판단을 위한 요인에서 가장 중요성이 낮았으나, 중요하다고 응답한 사람도 38.8%였음
- 기자가 ‘오보인 경우 빠르게 기사를 수정하고 공개하는지’와 ‘일반 대중이 생각하는 문제를 이해하고 있는지’가 기사 요인에서 중요한 것으로 나타남
- 언론사 요인에서는 과거 기록이 신뢰도 판단에 큰 영향을 미치는 것으로 나타남. 과거 오보 여부나 허위정보 작성 여부가 신뢰도 판단에 주요 요인이었음
- 타인에게 받는 요인은 다른 요인에 비해 중요성이 낮았으며, 많은 사람들이 보는 뉴스가 그중에서 상대적으로 가장 중요하다는 답변을 얻음

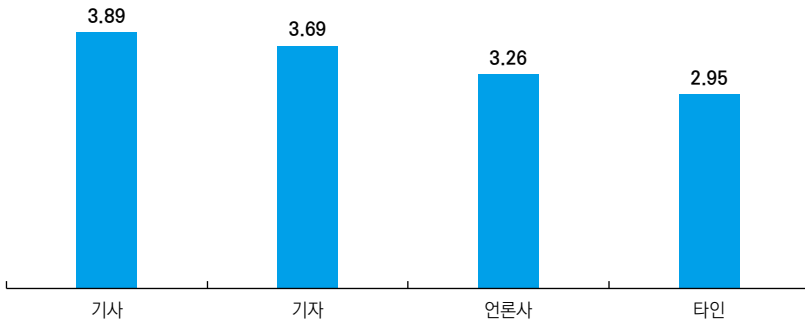
기계적으로 뉴스 신뢰도를 평가하기 위해서는 어떤 요인을 바탕으로 신뢰도를 판단하는지 확인하는 것이 필요하다. 뉴스 신뢰도를 기계적으로 평가할 때 사용하는 요인의 일반적 사례로 ‘기사의 길이’라든지 ‘언론사의 규모’와 같은 것이 있다. 하지만 이러한 요인은 실제 신뢰도를 측정하기 위한 다양한 간접적인 측정 방식 중 일부이다. 본 설문에서는 뉴스 이용자가 실제 뉴스를 접했을 때 어떤 요인을 보고 뉴스 신뢰도를 평가하는지 확인하고자 했다. 이를 위해 사람들이 뉴스와 관련된 신뢰도를 판단할 때 어떤 요인이 중요하다고 생각하는지 물었다. 기존 문헌에서 뉴스 신뢰도를 판단하는 데 사용될 수 있다고 제시된 28가지 문항을 제시했다. 응답자들은 각 요인이 포털에서 뉴스 신뢰도를 판단하는 데 얼마나 중요하다고 생각하는지 답했다.

뉴스 이용자에게 묻은 28가지 요인을 기사, 기자, 언론사, 타인의 네 가지 범주로 분류하였다. 각 범주의 평균값을 살펴보았을 때 가장 중요성이 높은

범주는 기사 자체와 관련된 것으로 5점 만점 기준 평균 3.89점이었다. 두 번째로 중요한 것은 기자와 관련된 범주로 평균 3.69점이었다. 세 번째는 언론사에 관한 범주로 평균 3.26점이었으며, 타인과 관련된 요인은 뉴스 기사 신뢰도를 판단하는 범주 중에서 가장 낮은 중요성(평균 2.95점)을 가지는 것으로 나타났다.

〈그림 4-22〉 뉴스를 접했을 때 신뢰도 평가에 영향을 미치는 요인

(단위: 점, n=1,000)

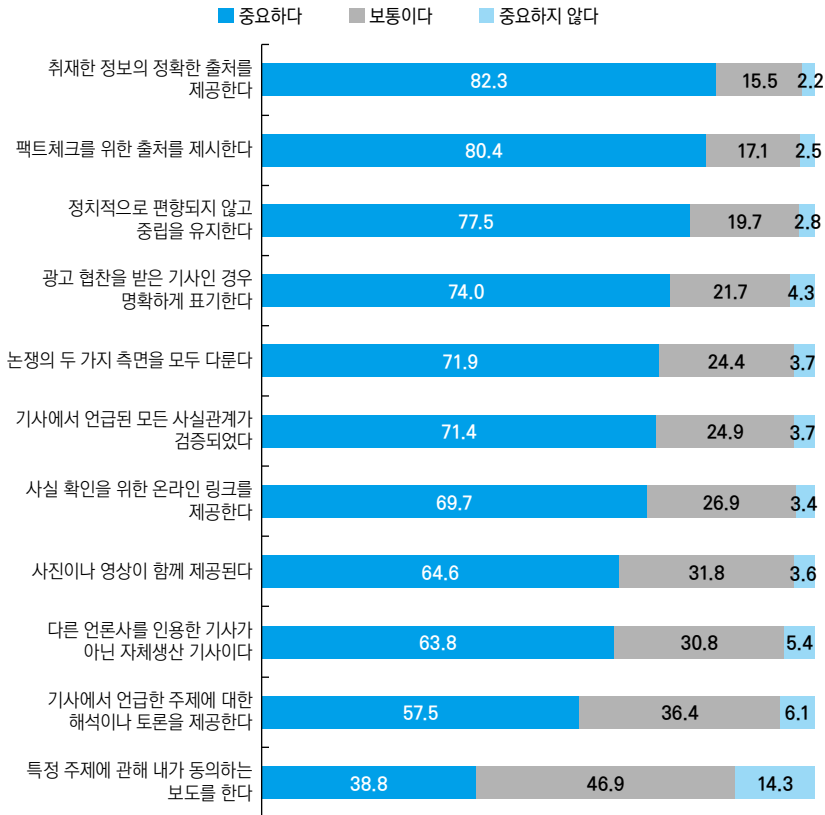


기사와 관련된 다양한 요인 중에서 중요하다는 답변이 가장 많은 것은 ‘취재한 정보의 정확한 출처를 제공한다.’로 82.3%가 중요하다고 응답했다. ‘팩트체크를 위한 출처를 제공한다.’가 80.4%로 두 번째였다. ‘정치적으로 편향되지 않고 중립을 유지한다.’가 77.5%, ‘광고 협찬을 받은 기사인 경우 명확하게 표기한다.’가 74.0%, ‘논쟁의 두 가지 측면을 다룬다.’가 71.9%, ‘기사에서 언급된 모든 사실관계가 검증되었다.’가 71.4%였다. 중요성이 높은 요인을 순서대로 보았을 때, 출처 확인이나 사실검증 관련된 기사의 객관성, 중립적인 보도와 관련된 공정성, 협찬 기사의 명확한 표현과 같은 투명성이 기사 신뢰도를 판단하는 주요한 요인임을 확인할 수 있었다. ‘온라인 링크 제공’, ‘사진이나 영상’은 중요하다는 응답이 상대적으로 낮았으며, ‘자체 생산 기사 여부’나 ‘주제에 관한 해석이나 토론 제공 여부’도 상대적으로 부차적 요인이었다. 흥미로운 사실은

‘특정 주제에 관해 내가 동의하는 보도를 한다.’라는 요인이었다. 신뢰도 판단에 중요하다고 답한 응답자는 가장 낮은 수준이었으나 38.8%가 신뢰하는 요인으로 동의했다. 즉, 대부분의 뉴스 이용자는 전통적인 신뢰도 평가 요인인 객관성, 공정성, 투명성을 통해 뉴스의 신뢰 여부를 확인하지만, 일부 응답자는 기사 신뢰도를 평가하는 요인으로 자신의 의견과 일치하는지를 통해 판단한다는 사실을 알 수 있었다.

〈그림 4-23〉 뉴스를 접했을 때 신뢰도 평가에 영향을 미치는 요인(기사)

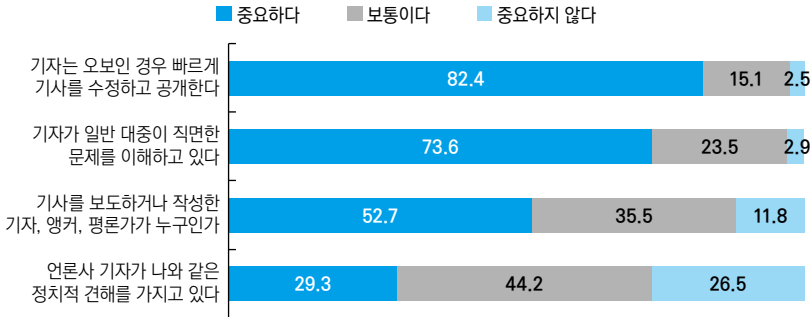
(단위: %, n=1,000)



기사를 통해 판단하는 신뢰도 여부는 기사와 관련된 요인보다는 상대적으로 중요성이 낮았다. 하지만 그중에서 ‘기사는 오보인 경우 빠르게 기사를 수정하고 공개한다.’라는 요인이 중요하다는 응답은 82.4%였으며, ‘기자가 일반 대중이 직면한 문제를 이해하고 있다.’라는 응답은 73.6%였다. 하지만 ‘기사를 보도 하거나 작성한 언론인이 누구인지’라든가 ‘기자가 나와 같은 정치적 견해를 가지고 있는지’ 여부가 중요하다는 응답은 각각 52.7%, 29.3%로 중요성이 상대적으로 낮았다.

〈그림 4-24〉 뉴스를 접했을 때 신뢰도 평가에 영향을 미치는 요인(기사)

(단위: %, n=1,000)

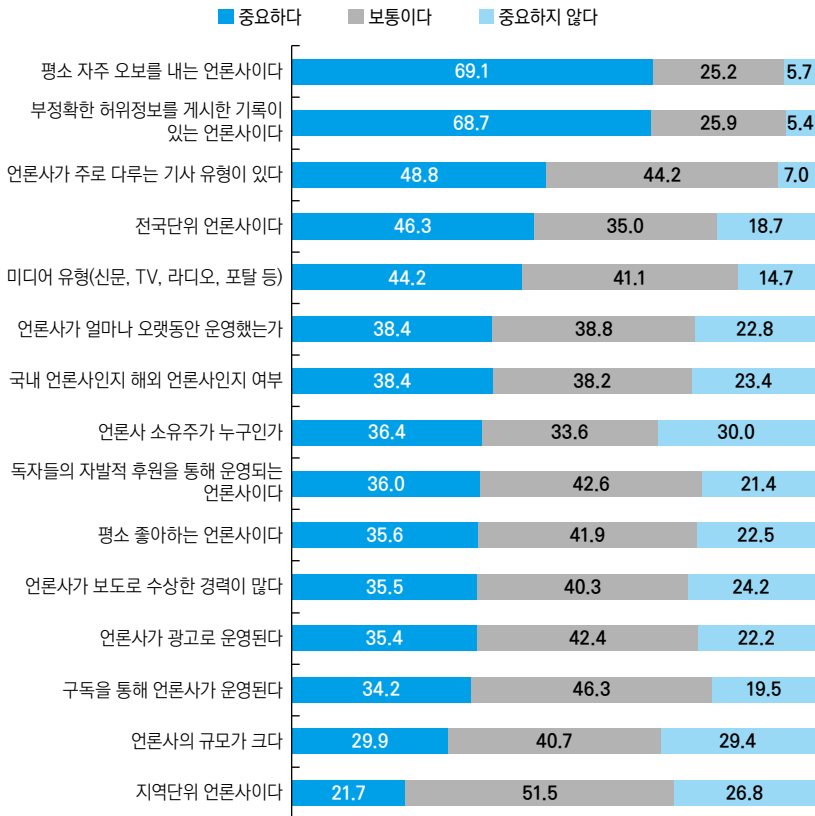


언론사를 통해 판단하는 기사 신뢰도 요인 중 높은 중요성이 있는 요인은 ‘평소 자주 오보를 내는 언론사이다.’와 ‘부정확한 허위정보를 제시한 기록이 있는 언론사이다.’로, 이를 중요하다고 응답한 비율은 각각 69.1%, 68.7%였다. 나머지 요인이 중요하다는 응답은 절반을 넘지 못했다. ‘언론사가 주로 다루는 기사 유형이 있다.’는 48.8%, ‘전국단위 언론사이다.’는 46.3%, ‘미디어 유형’ 44.2%, ‘언론사가 얼마나 오랫동안 운영했는지’는 38.4%, ‘국내 언론사 여부’는 38.4%, ‘언론사 소유주’는 36.4%가 중요하다고 답했다. 나머지 요인인 언론사의 ‘수익모델’, ‘수상 경력’, ‘규모’와 같은 것들이 뉴스 신뢰도 판단에 중요하다는 응답은 높지 않았다. 주목할 만한 사실은 ‘언론사 소유주’와 같은 요인이 중요하지 않다는 응답도 30%였다는 점이다. 또한, 언론사가 광고로 운영되는지, 구독

을 통해 운영되는지, 자발적 후원을 통해 운영되는지가 신뢰도 평가에 중요하지 않다는 응답도 20% 전후인 것을 확인할 수 있었다. 전통적으로 저널리즘에서 언론이 자본으로부터 독립되어야 한다는 점을 중요하게 지적했으나, 뉴스 이용자로서는 기사가 믿을 만한지 판단하는 요인으로서 중요성이 낮다고 인지하는 것으로 나타났다.

〈그림 4-25〉 뉴스를 접했을 때 신뢰도 평가에 영향을 미치는 요인(언론사)

(단위: %, n=1,000)

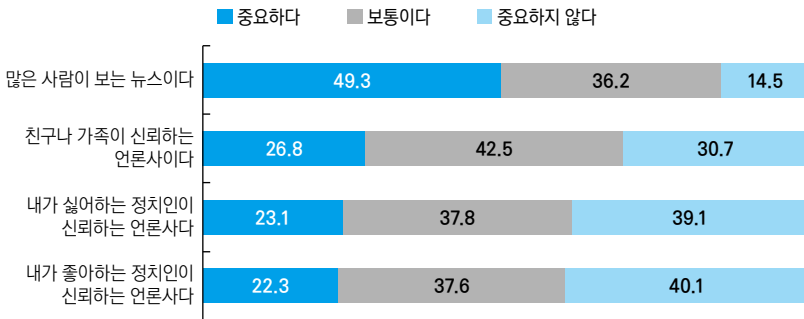


기사의 신뢰도 판단에 타인이 영향을 미칠 수 있는 부분도 있다. 본 설문에서는 타인이 신뢰도 판단에 영향을 미칠 수 있는 요인 네 가지를 물었다. 그중에서 ‘많은 사람들이 보는 뉴스’가 신뢰도 평가에서 중요한 요인이라는 응답이 49.3%로 가장 높았다. 나머지 응답은 모두 중요하지 않다는 답변이 중요하다는 응답보다 더 높았다. ‘친구나 가족이 신뢰하는 언론사이다.’라는 응답은 26.8%, ‘내가 싫어하는 정치인이 신뢰하는 언론사이다.’라는 응답은 23.1%, ‘내가 좋아하는 정치인이 신뢰하는 언론사이다.’라는 응답은 22.3%만이 중요하다고 답했다.

전반적으로 신뢰도 평가에 영향을 미치는 주요 요인은 기사 요인으로 대부분 응답자가 기사에서 발견할 수 있는 객관성, 공정성, 투명성이라든지 언론사가 쌓아온 신뢰도와 같은 전통적인 신뢰도 평가 요인이 중요하다고 판단하는 것으로 나타났다. 반면, 언론사의 수익모델과 같은 부분이 신뢰도 판단에 중요하다는 답변은 상대적으로 낮았다. 이러한 부분은 실제 언론사의 수익모델이 저널리즘 신뢰도와 영향이 없다기보다는 일반 대중이 개별 언론사의 수익모델을 파악하는 것이 어렵기 때문일 수도 있다. 다음에서는 이러한 요인을 포함하여 개별 기사 및 뉴스 배열에 영향을 미치는 요인에 관한 실험을 하였다.

〈그림 4-26〉 뉴스를 접했을 때 신뢰도 평가에 영향을 미치는 요인(타인)

(단위: %, n=1,000)



4) 기사 신뢰도 평가 요인 실험

- 신뢰 지표로 기사 작성시간, 기사 수정시간, 기사 수정 내역, 기사 분류정보, 취재원 정보, 기자 사진, 기자 이메일, 기자 이력에 관한 내용을 포함한 기사는 언론사에 대한 일반적 평가에는 크게 유의미한 영향을 주지 않음
- 신뢰 지표를 포함한 기사는 언론사 취재와 기자에 관한 평가에서는 통계적으로 유의하게 높은 평가를 이끌었으며, 언론사에 관한 만족도, 추천 의사, 유료구독 의사에 신뢰 지표가 직접 영향을 주지는 못했음
- 이용자가 인지하기 쉬운 지표는 기자 사진이었으나, 실제 기자 사진이 뉴스 신뢰도에 영향을 미치는 중요한 요인이라는 응답은 40.2%로 낮은 편으로 나타남
- 각 신뢰 지표 인지 여부에 따라 언론사의 믿을 만한 정도를 살펴본 결과에 따르면 ‘기자 이력’, ‘기사 분류정보’, ‘기사 수정시간’, ‘기사 수정 내역’이 신뢰를 평가하는 주요 요인이었음
- 일반적으로 중요하다고 생각하는 신뢰 요인은 ‘기사 수정내역’, ‘취재원 정보’, ‘기사 수정시간’ 순으로 나타남

뉴스 신뢰도가 낮은 상황에서 신뢰할 수 없는 정보의 확산은 점점 심해지고 있다. 이러한 상황에서 대중들은 기사에서 드러나는 다양한 단서를 통해 뉴스가 신뢰할 수 있는지를 판단해야 한다. 본 연구는 다양한 신뢰 지표를 포함한 기사와 포함하지 않은 기사를 1,000명의 응답자에게 보여주었다. 실험에 사용한 신뢰 지표는 기사 작성시간, 기사 수정시간, 기사 수정 내역, 기사 분류정보, 취재원 정보, 기자 사진, 기자 이메일, 기자 이력에 관한 내용을 포함했다. 이러한 실험 결과는 언론사는 대중의 신뢰를 높이는 조치를 할 수 있을 뿐만 아니라 기사의 신뢰도를 기계적으로 평가하는 알고리즘을 고려하기 위한 사전 연구로 활용할 수 있을 것이다.

〈그림 4-27〉 신뢰 지표 실험(왼쪽=신뢰 지표 포함, 오른쪽=신뢰 지표 미포함)

ACC

주요뉴스 정치 경제 사회 생활 세계 지역방송 생중계

ACC

서울 어제 51명 확진...22일 만에 50명대로 줄어

입력 2020.09.05. 오후 7:51 | 수정 2020.09.05. 오후 8:14

오세욱 기자 karpah@acc.co.kr

이 기사는 언론사에서 사회 섹션으로 분류했습니다.



요양원에 임시한 선별진료소... 밤샘하긴 일러
2020.9.5 karpah@acc.co.kr

서울의 신종 코로나바이러스 감염증(코로나19) 확진자가 지난 4일 하루 동안 51명 늘었다고 서울시가 5일 밝혔다.

이는 지난달 13일(32명) 이래 22일 만에 가장 적은 수치다. 서울의 신규 확진자는 지난달 26일 154명으로 정점을 찍은 뒤 점차 줄어들어 지난 2일과 3일에는 각각 69명, 68명을 기록했다.

누적 확진자는 4천251명으로 집계됐다. 확산세는 누그러졌지만 종교·체육시설과 의료기관을 중심으로 산발적 집단감염이 계속됐다.

중앙사고수습본부에 따르면 전날 기준 비어 있는 서울의 중증환자 병상은 7개로 하루 전보다 2개 늘었다. 이 가운데 장비와 인력이 완비돼 확진자가 즉시 입원할 수 있는 병상 역시 2개에서 5개로 증가했다.

병상을 공동 운영하는 경기도·인천시를 합하면 10개가 비어 있고 이 중 7개를 즉시 사용할 수 있는 것으로 집계됐다.

서울의 감염병 전담병원 병상 가동률은 71.8%로 하루 전 72.3%에서 소폭 하락했다.

뉴스 취재원 정보

- 서울시 언론담당관
- 중앙사고수습본부 본부장

수정 내역

- 확진자 숫자 변동으로 인한 기사 수정

ACC
오세욱 기자 >
ACC 뉴스 사회부 기자입니다.

ACC

주요뉴스 정치 경제 사회 생활 세계 지역방송 생중계

ACC

서울 어제 51명 확진...22일 만에 50명대로 줄어

윤라민 이슈 >



요양원에 임시한 선별진료소... 밤샘하긴 일러

서울의 신종 코로나바이러스 감염증(코로나19) 확진자가 지난 4일 하루 동안 51명 늘었다고 서울시가 5일 밝혔다.

이는 지난달 13일(32명) 이래 22일 만에 가장 적은 수치다. 서울의 신규 확진자는 지난달 26일 154명으로 정점을 찍은 뒤 점차 줄어들어 지난 2일과 3일에는 각각 69명, 68명을 기록했다.

누적 확진자는 4천251명으로 집계됐다. 확산세는 누그러졌지만 종교·체육시설과 의료기관을 중심으로 산발적 집단감염이 계속됐다.

중앙사고수습본부에 따르면 전날 기준 비어 있는 서울의 중증환자 병상은 7개로 하루 전보다 2개 늘었다. 이 가운데 장비와 인력이 완비돼 확진자가 즉시 입원할 수 있는 병상 역시 2개에서 5개로 증가했다.

병상을 공동 운영하는 경기도·인천시를 합하면 10개가 비어 있고 이 중 7개를 즉시 사용할 수 있는 것으로 집계됐다.

서울의 감염병 전담병원 병상 가동률은 71.8%로 하루 전 72.3%에서 소폭 하락했다.

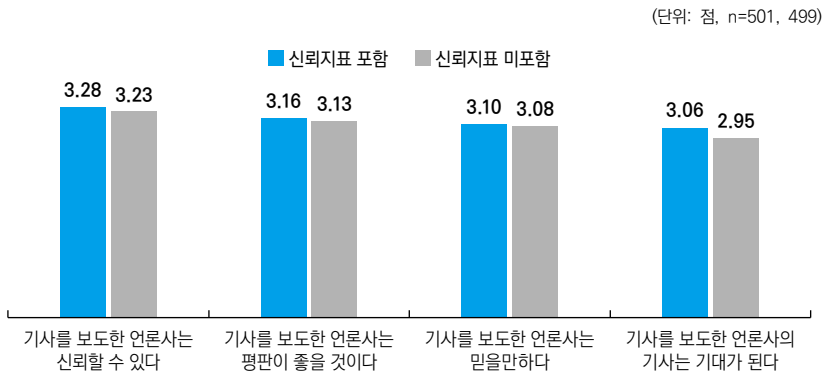
ACC
윤라민 뉴스 >

신뢰 지표 중 기사와 관련된 내용은 기사 입력시간과 수정시간을 모두 포함하고 있으며, 기사 수정 내역에서 어떤 이유로 기사를 수정했는지 기사 하단에 별도로 밝히고 있다. 기사 분류정보는 기사 상단에 파란색 글씨로 “이 기사는 언론사에서 사회 섹션으로 분류했습니다.”라는 문구를 포함했다. 출처는 기사에서 인용한 정보원이 누구인지에 대해 별도로 출처를 포함하였다. 기사와 관련된

정보로 기자가 속한 언론사, 기자 사진, 기자 이메일이 기사 상단과 하단에 포함되었다. 반면 신뢰 지표를 포함하지 않은 기사의 경우 기자 명이 익명의 ‘온라인 이슈팀’으로 표기되고 다른 정보를 하나도 포함하지 않았다. 그 외 기사 내용이나 기사에 포함된 사진은 두 기사가 같게 구성했다.

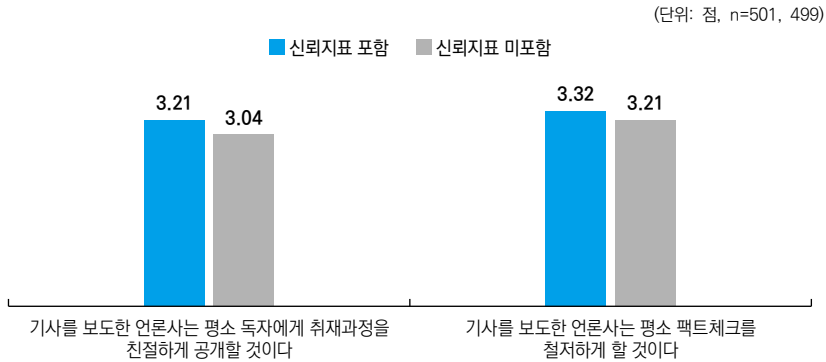
기사를 본 후 응답자는 다양한 언론사 평가 질문에 답변했다. 언론사에 관한 일반적 평가로 ‘언론사를 신뢰할 수 있는지’, ‘언론사 평판이 좋다고 생각하는지’, ‘언론사는 믿을 만한지’, ‘언론사의 기사가 기대가 되는지’ 네 가지를 물었다. ‘기사를 보도한 언론사는 신뢰할 수 있는지’에 관한 질문에 신뢰 지표 포함 기사의 평균은 3.28점, 미포함 기사의 평균은 3.23점으로 통계적으로 유의한 차이를 보이지 않았다($t = 0.99, p = n.s.$). ‘기사를 보도한 언론사는 평판이 좋을 것이다.’라는 질문은 신뢰 지표 포함 기사 평균 3.16점, 미포함 기사 평균은 3.13점으로 통계적으로 유의하지 않았다($t = 0.44, p = n.s.$). ‘기사를 보도한 언론사는 믿을 만하다.’라는 질문의 신뢰 지표 포함 평균은 3.10점, 미포함 평균은 3.08점으로 차이를 보이지 않았다($t = 0.57, p = n.s.$). 언론사에 관한 일반적 평가 항목 중 유일하게 통계적으로 차이를 보인 것은 ‘기사를 보도한 언론사의 기사는 기대가 된다.’라는 항목으로 신뢰 지표 포함 평균 3.06점, 미포함 평균 2.95로 통계적으로 유의한 차이를 보였다($t = 2.08, p < .05$).

〈그림 4-28〉 기사 작성 언론사에 관한 일반적 평가

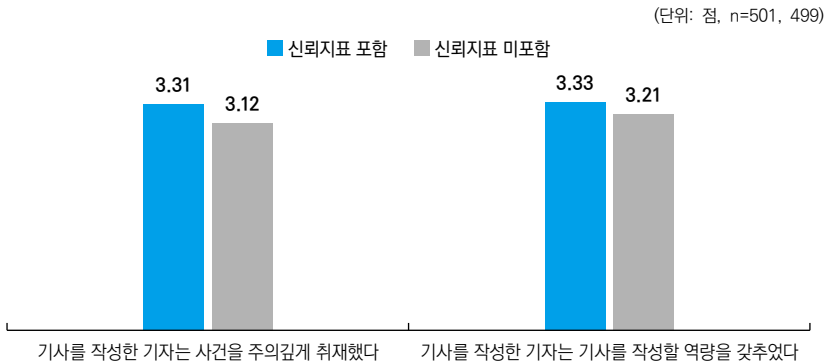


언론사 취재와 관련된 평가로 언론사의 ‘평소 취재 과정 공개 여부’, ‘팩트체크 여부’에 관한 인식을 물었다. ‘기사를 보도한 언론사는 평소 독자에게 취재 과정을 친절하게 공개할 것’이라고 생각하는지에 신뢰 지표 포함 기사는 평균 3.21점, 미포함 기사는 평균 3.04점으로 통계적으로 유의한 차이를 보였다($t = 3.39$, $p < .001$). ‘기사를 보도한 언론사는 평소 팩트체크를 철저하게 할 것’이라고 생각하는지에 신뢰 지표 포함 기사는 평균 3.32점, 미포함 기사는 평균 3.21점으로 통계적으로 유의했다($t = 2.20$, $p < .05$). 이러한 결과는 신뢰 지표가 언론사 취재 과정에 관한 인식에는 유의한 영향을 준다는 사실을 확인할 수 있었다.

〈그림 4-29〉 언론사의 취재에 관한 평가

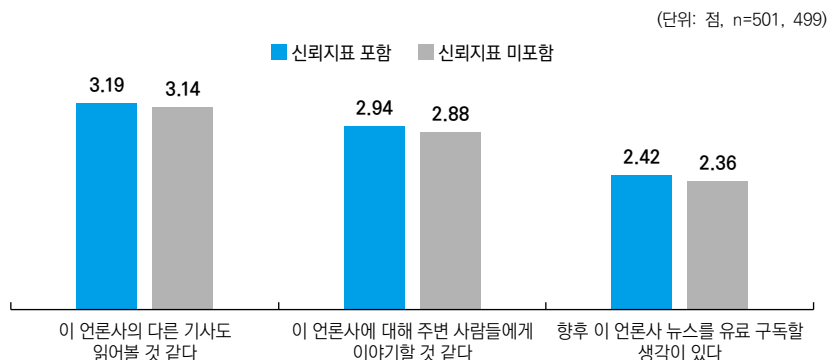


〈그림 4-30〉 기자에 관한 평가



기자에 관한 평가는 ‘사건 취재에 관한 인식’과 ‘기자 역량’에 관해 물었다. ‘기사를 작성한 기자는 사건을 주의 깊게 취재했다.’라는 질문은 신뢰 지표 포함 평균 3.31점, 미포함 3.12점으로 통계적으로 유의한 차이를 보였다($t = 3.51$, $p < .001$). ‘기사를 작성한 기자는 기사를 작성할 역량을 갖추었다.’라는 질문은 신뢰 지표 포함 평균 3.33점, 미포함 3.21점으로 통계적으로 유의했다($t = 2.35$, $p < .05$). 신뢰 지표가 포함되었을 때 기자에 대한 평가가 포함되지 않았을 때와 비교하여 전반적으로 높다는 사실을 확인할 수 있었다.

〈그림 4-31〉 언론사 만족도 및 구독 의사 평가

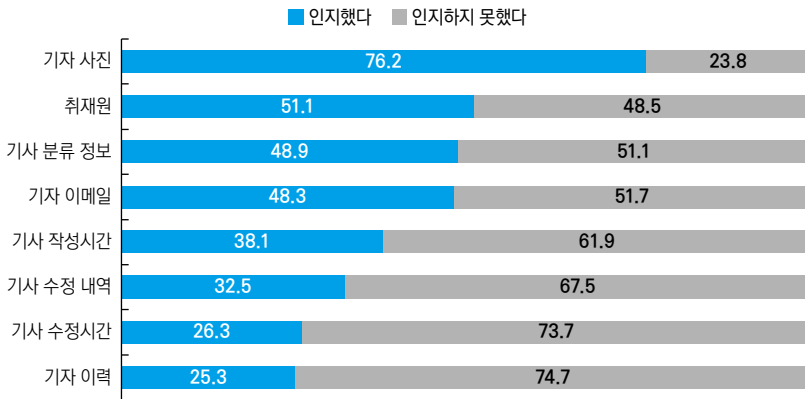


언론사에 대한 만족도, 추천 의사, 유료구독 의사에 관한 질문에서 신뢰 지표 포함 여부는 통계적으로 유의한 차이를 보이지는 않았다. ‘이 언론사의 다른 기사도 읽어볼 것 같다.’라는 질문은 신뢰 지표 포함 평균 3.19점, 미포함 평균 3.14점이었다. ‘이 언론사에 대해 주변 사람들에게 이야기할 것 같다.’라는 질문은 신뢰 지표 포함 평균 2.94점, 미포함 평균 2.88점이었다. ‘향후 이 언론사 뉴스를 유료 구독할 생각이 있다.’라는 질문은 신뢰 지표 포함 평균 2.42점, 미포함 평균 2.36점이었다. 뉴스 이용자의 행위 의사와 관련된 질문에서는 신뢰 지표 포함 기사가 점수는 약간 높았으나 통계적으로는 유의하지 않은 것으로 나타났다.

실험에 사용된 신뢰 요인 중 응답자가 어떤 요인을 인지했는지를 신뢰 요인이 포함된 기사를 본 501명을 대상으로 물었다. 가장 많은 사람이 인지했다고 답한 요인은 기자 사진으로 76.2% 응답자가 인지했다고 답했다. 취재원 정보를 확인했다는 사람은 51.5%였다. 나머지 요인을 인지한 응답자는 절반 이하로 낮은 수준이었다. 기사 분류정보가 48.9%, 기자 이메일이 48.3%, 기사 작성시간이 38.1%, 기사 수정 내역이 32.5%, 기사 수정시간이 26.3%, 기자 이력이 25.3%였다. 실험 설계의 제약으로 인해 해당 내용을 구체적으로 인지하지 못한 응답자의 비율이 높은 편이었다.

〈그림 4-32〉 신뢰 요인의 인지 여부

(단위: %, n=501)

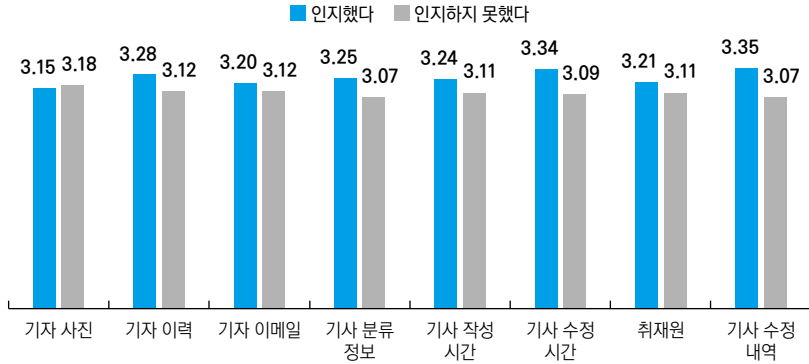


본 실험에 사용한 신뢰 요인은 트러스트 프로젝트(The Trust Project)에서 사용한 투명성 관련 요인의 일부를 활용하여 구성한 것이다. 기존 연구는 이러한 신뢰 지표가 포함된 뉴스를 본 이용자는 뉴스 조직의 평판에 관해 더 높은 평가를 하는 것으로 알려졌다. 하지만 본 실험은 개별 요인에 대한 효과를 검증하는 것이 아닌 전반적인 차이만을 살펴보는 간략한 실험 설계를 사용했다. 또한, 실제 트러스트 프로젝트에서 사용된 것보다 간단한 요인을 사용했다. 예를 들면, 텍사스 오스틴 대학의 미디어인게이지먼트센터(Center for Media

Engagement)는 기사에 대한 설명 상자에 취재와 관련된 뒷이야기(behind story)를 구체적으로 포함하는 방식을 사용했다.

〈그림 4-33〉 신뢰 요인의 인지 여부에 따른 믿을 만한 정도 차이

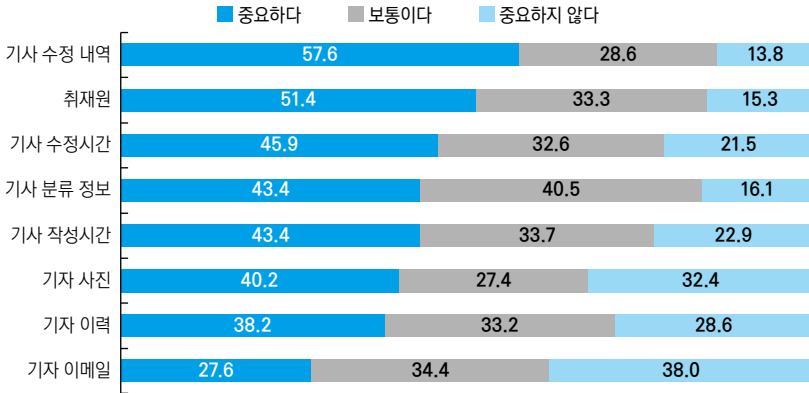
(단위: %, n=501)



기사를 보도한 언론사가 믿을 만하다고 생각하는지를 각 신뢰 요인의 인지 여부에 따라 나누어 살펴본 결과에 따르면 몇 가지 요인은 언론사가 믿을 만하다고 판단하는 데 통계적으로 유의한 영향을 미친다는 사실을 확인할 수 있었다. 기자 이력은 인지한 응답자가 평균 3.28점, 미인지 응답자가 3.12점($t = 2.05$, $p < .05$)이었고, 기사 분류정보는 인지한 응답자가 평균 3.25점, 미인지 응답자가 3.07점($t = 2.78$, $p < .01$), 기사 수정시간은 인지한 응답자 평균 3.34, 미인지 응답자 평균 3.09($t = 3.31$, $p < .001$), 기사 수정 내역은 인지한 응답자 평균 3.35, 미인지 응답자 평균 3.07($t = 4.06$, $p < .001$)이었다. 즉, 기자 이력, 기사 분류정보, 기사 수정시간, 기사 수정 내역은 인지한 이용자와 인지하지 못한 이용자 사이 통계적으로 유의한 차이를 보였다.

〈그림 4-34〉 신뢰 요인이 뉴스 신뢰도에 미치는 영향력

(단위: %, n=1,000)



실험에 사용한 신뢰 요인이 뉴스 신뢰도에 미치는 중요성에 관한 생각을 물었다. 응답자가 가장 중요하다고 생각하는 신뢰 요인은 ‘기사 수정 내역’으로 57.6%의 응답자가 중요하다고 답했다. ‘취재원에 관한 정보’가 51.4%, ‘기사 수정시간’이 45.9%였다. 다음으로 ‘기사 분류정보’와 ‘기사 작성시간’이 각각 43.4%, ‘기자 사진’ 40.2%, ‘기자 이력’ 38.2%, ‘기자 이메일’이 27.6%였다. 인지 여부에서 가장 높은 인지도를 보이는 것은 기자 사진이었으나 ‘기자 사진’이 중요하다고 답한 응답자는 40.2%에 그쳤으며, 뉴스 트러스트 위원회에서 주요한 요인으로 사용했던 ‘기자 이메일’이 중요하다는 응답자는 27.6%로 중요하지 않다고 판단하는 응답자가 38.0%로 더 많았다.

5) 배열 신뢰도 평가 실험

- 뉴스 배열에서 단순한 신뢰도 인증표시보다는 알고리즘 신뢰도 평가 점수가 포함되었을 때 뉴스 이용자가 느끼는 배열 신뢰도가 더 높았음. 다만 점수 포함 여부는 공정성, 다양성, 정확성, 심층성, 투명성, 객관성, 화제성과 같은 다른 요인에 의미 있는 영향을 미치지 못함

- 뉴스 배열에서 언론사명 포함 여부와 점수 배열방식(내림차순 vs. 무작위)의 영향력은 배열 신뢰도에는 영향을 주지 못함. 하지만 점수를 내림차순으로 배열하는 것이 뉴스 이용자에게 뉴스 배열이 어느 한쪽으로 치우치지 않았다는 느낌을 주는 것으로 나타남

〈그림 4-35〉 뉴스 배열 신뢰도 실험(점수 포함 vs. 미포함)

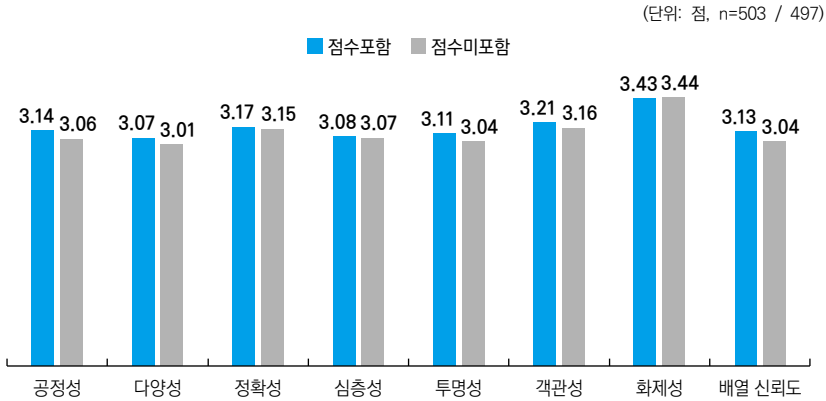
(단위: %, n=503/497)

| | |
|--|---|
| <p>코로나 치료 '인포데믹' 퍼뜨리는 지도자들 경향신문 : 09월 07일 15:19</p> <p>56</p> | <p>코로나 치료 '인포데믹' 퍼뜨리는 지도자들 경향신문 : 09월 07일 15:19</p> <p>56</p> |
| <p>"소금물 소독은 인포데믹...감염병보다 위험" 경향신문 : 09월 07일 15:55</p> <p>33</p> | <p>"소금물 소독은 인포데믹...감염병보다 위험" 경향신문 : 09월 07일 15:55</p> <p>33</p> |
| <p>'메탄을 뿌려서 소독?...인포데믹' 피하러면? JBK : 09월 07일 15:35</p> <p>70</p> | <p>'메탄을 뿌려서 소독?...인포데믹' 피하러면? JBK : 09월 07일 15:35</p> <p>70</p> |
| <p>"5G로 전염된다?...코로나19 음모론 기승" 부경일보 : 09월 07일 15:55</p> <p>34</p> | <p>"5G로 전염된다?...코로나19 음모론 기승" 부경일보 : 09월 07일 15:55</p> <p>34</p> |
| <p>코로나19, 악의적인 '교란정보'가 가장 큰 문제 미디어오늘 : 09월 07일 15:20</p> <p>85</p> | <p>코로나19, 악의적인 '교란정보'가 가장 큰 문제 미디어오늘 : 09월 07일 15:20</p> <p>85</p> |
| <p>'인포데믹'으로 변이한 코로나19...음모론 몰살 LBS : 09월 07일 15:25</p> <p>67</p> | <p>'인포데믹'으로 변이한 코로나19...음모론 몰살 LBS : 09월 07일 15:25</p> <p>67</p> |
| <p>뉴스 트러스트 프로젝트는 뉴스 기사에 대한 품질을 기계적으로 평가하는 기준을 마련하여 해당 기사가 신뢰할 만한지를 확인받았다는 마크와 함께 0점부터 100점까지 점수로 표기합니다. 다음은 팬데믹과 관련된 주제를 알고리즘으로 배열한 결과입니다. 뉴스 배열을 확인 후 다음 질문에 답변해주시오.</p> | <p>뉴스 트러스트 프로젝트는 뉴스 기사에 대한 품질을 기계적으로 평가하는 기준을 마련하여 해당 기사가 신뢰할 만한지를 확인받았다는 마크를 제공합니다. 다음은 팬데믹과 관련된 주제를 알고리즘으로 배열한 결과입니다. 뉴스 배열을 확인 후 다음 질문에 답변해주시오.</p> |

기계적으로 개별 기사 신뢰를 평가한 후 뉴스 배열에 기사 평가를 도입했을 때, 뉴스 이용자들이 뉴스 배열에 대해 느끼는 신뢰도를 확인하기 위해 두 가지 실험을 하였다. 첫 번째 실험은 뉴스 배열에 신뢰도 평가 점수가 포함되었을 때 뉴스 배열과 관련하여 이용자가 느끼는 공정성, 다양성, 정확성, 심층성, 투명성, 객관성, 화제성을 살펴보고 편집 결과에 대한 신뢰성을 물었다. 실험 참가자는 뉴스 신뢰도 평가 점수가 포함된 배열과 뉴스 트러스트 마크가 포함된

배열 두 가지에 임의로 할당하였다(그림 4-35). 참가자가 뉴스 배열을 확인하기 전 해당 배열에 관해 제시된 안내문을 읽었다. 공통점은 뉴스 트러스트 프로젝트는 뉴스 기사에 대한 품질을 기계적으로 평가하는 기준을 마련하여 해당 기사가 신뢰할 만한지를 나타낸다고 한 것이고, 차이점은 0점부터 100점까지 점수가 포함되는지 또는 신뢰할 만하다는 마크만 제시하는지였다.

〈그림 4-36〉 뉴스 배열 신뢰도 실험 결과



결과에 따르면 공정성, 다양성, 정확성, 심층성, 투명성, 객관성 화제성 모두 신뢰도 평가 점수가 포함되었을 때 더 높은 평균 점수를 얻었다. 하지만 각 평가 기준의 점수 차이는 통계적으로 유의하지 않았다. 다만 배열 신뢰도의 경우만 통계적으로 유의한 차이를 보였다. 점수가 포함된 기사에 관한 신뢰도 점수는 평균 3.13점, 미포함된 기사에 관한 신뢰도 점수는 평균 3.04점이었다($t = 2.03$, $p < .05$). 이는 뉴스 신뢰도 점수를 포함하는 배열이 다양한 평가 기준의 차이를 가져오지는 못하지만, 뉴스 이용자에게 배열 자체에 관한 신뢰도를 높일 수 있다는 사실을 보여준다.

〈그림 4-37〉 뉴스 배열 신뢰도 실험(언론사명, 점수 배열방식)

(단위: %, n=252/241/250/257)

| | | | |
|---|----|--|----|
| "한국말 못하면 왜 욕하고 바보 취급하는지"...한국 인증자별 심각 서경신문 09월 07일 15:19 | 76 | "한국말 못하면 왜 욕하고 바보 취급하는지"...한국 인증자별 심각 09월 07일 15:19 | 76 |
| 인권감수성 높은 '코로나의 역설'...88%가 "차별금지법 찬성" 상경일보 09월 07일 15:55 | 62 | 인권감수성 높은 '코로나의 역설'...88%가 "차별금지법 찬성" 09월 07일 15:55 | 62 |
| '코로나 차별혐오'에 국민 90% "나도 차별 당할 수 있다" 부대신문 09월 07일 15:35 | 53 | '코로나 차별혐오'에 국민 90% "나도 차별 당할 수 있다" 09월 07일 15:35 | 53 |
| 귀화선수도 '인증자별' 파문...혐오 만연하는 까닭은 WIS 09월 07일 15:55 | 41 | 귀화선수도 '인증자별' 파문...혐오 만연하는 까닭은 09월 07일 15:55 | 41 |
| "인증자별을 남의 나라 문제로 보는 안이한 태도 버려야" 미디어인 09월 07일 15:20 | 16 | "인증자별을 남의 나라 문제로 보는 안이한 태도 버려야" 09월 07일 15:20 | 16 |
| "네 나라로 돌아가라" 멈추지 않는 동양인 혐오 JBC 09월 07일 15:25 | 13 | "네 나라로 돌아가라" 멈추지 않는 동양인 혐오 09월 07일 15:25 | 13 |

언론사 포함 / 점수 내림차순

언론사 미포함 / 점수 내림차순

| | | | |
|--|----|--|----|
| "인증자별을 남의 나라 문제로 보는 안이한 태도 버려야" 서경신문 09월 07일 15:20 | 16 | "인증자별을 남의 나라 문제로 보는 안이한 태도 버려야" 09월 07일 15:20 | 16 |
| 인권감수성 높은 '코로나의 역설'...88%가 "차별금지법 찬성" 상경일보 09월 07일 15:55 | 62 | 인권감수성 높은 '코로나의 역설'...88%가 "차별금지법 찬성" 09월 07일 15:55 | 62 |
| "네 나라로 돌아가라" 멈추지 않는 동양인 혐오 부대신문 09월 07일 15:25 | 13 | "네 나라로 돌아가라" 멈추지 않는 동양인 혐오 09월 07일 15:25 | 13 |
| 귀화선수도 '인증자별' 파문...혐오 만연하는 까닭은 WIS 09월 07일 15:55 | 41 | 귀화선수도 '인증자별' 파문...혐오 만연하는 까닭은 09월 07일 15:55 | 41 |
| "한국말 못하면 왜 욕하고 바보 취급하는지"...한국 인증자별 심각 미디어 09월 07일 15:19 | 76 | "한국말 못하면 왜 욕하고 바보 취급하는지"...한국 인증자별 심각 09월 07일 15:19 | 76 |
| '코로나 차별혐오'에 국민 90% "나도 차별 당할 수 있다" JBC 09월 07일 15:35 | 53 | '코로나 차별혐오'에 국민 90% "나도 차별 당할 수 있다" 09월 07일 15:35 | 53 |

언론사 포함 / 점수 무작위

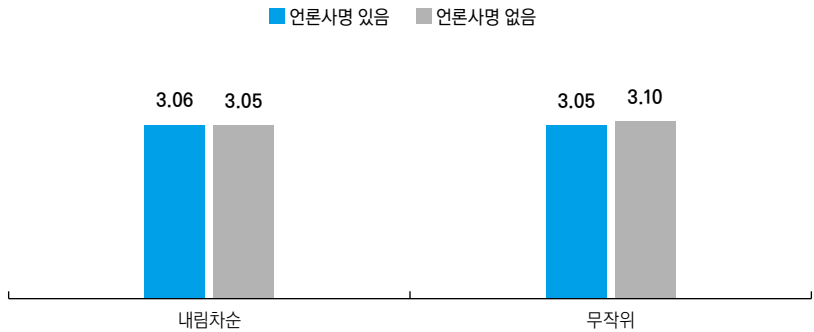
언론사 미포함 / 점수 무작위

두 번째 실험은 언론사명과 점수 배열방식이 뉴스 배열 신뢰도에 미치는 영향을 살펴보고자 했다. 언론사명이 포함된 기사와 포함되지 않은 기사, 점수에 따른 배열방식이 높은 점수부터 내림차순인 경우와 무작위인 경우로 나누어 전체 네 가지 뉴스 배열방식이 제시되었다. 실험 참가자는 각 조건에 따른 네 가지

뉴스 배열에 임의로 할당되어 기사 배열방식을 확인한 후, 이전과 같이 배열 평가 기준과 신뢰도에 관해 답했다. 결과에 따르면 뉴스 배열방식은 신뢰도에는 아무 영향을 주지 못한다는 사실을 확인할 수 있었다. 내림차순의 경우 언론사명 포함 평균 3.06점, 미포함 평균 3.05점으로 거의 차이가 나지 않았다. 무작위 순의 경우 언론사명 포함 배열 평균 3.05점, 미포함 평균 3.10점으로 통계적으로 유의하지 않았다.

〈그림 4-38〉 뉴스 배열 신뢰도(언론사명, 점수 배열방식)

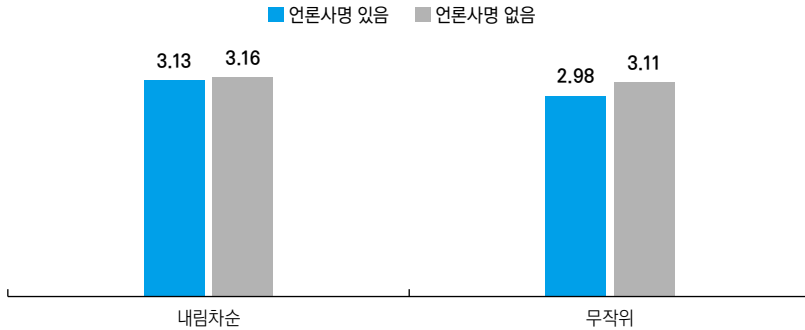
(단위: 점)



평가 기준에서도 점수 배열방식이나 언론사명 유무가 유의한 영향을 주지는 못했으나, 공정성과 관련된 평가 기준에서만 점수 배열방식이 영향을 미쳤다. ‘뉴스 배열이 어느 한쪽으로 치우치지 않았다.’라는 공정성과 관련된 질문에 내림차순이 무작위보다 높은 평균 점수를 기록하였으며 통계적으로도 유의한 차이를 보였다($F = 3.90, p < .05$). 즉, 기계적 신뢰도 평가가 포함되었을 때, 뉴스 이용자는 점수가 높은 순서에 따른 뉴스 배열을 공정하다고 느끼는 것으로 볼 수 있다.

〈그림 4-39〉 뉴스 배열 공정성(언론사명, 점수 배열방식)

(단위: 점)



6) 알고리즘 리터러시

- 미디어 리터러시에서 접근성에 관한 질문은 응답자의 70.9%가 자신 있다고 답했지만, 알고리즘의 영향력(36.4%)이나 가짜 뉴스 인식(27.9%)은 낮은 수준을 보였음
- 실제 기사 작성 과정을 경험해 본 사람은 10명 중 1명 수준으로 뉴스 리터러시 교육 여지가 많다는 사실을 확인할 수 있음
- 응답자는 미디어의 비판적 평가 방식을 이해할 필요가 있다는 데 66.2%가 공감했으며, 허위정보·오정보 인식 방법 이해가 중요하다는 데 51.9%가 동의했음
- 포털 뉴스 편집이 사회적으로 미치는 영향력은 긍정과 부정이 유사한 수준으로 나타났으나, 알고리즘 규제(74.7%), 포털 뉴스 규제(67.2%)를 지지하는 응답자가 다수였음

(1) 뉴스 리터러시

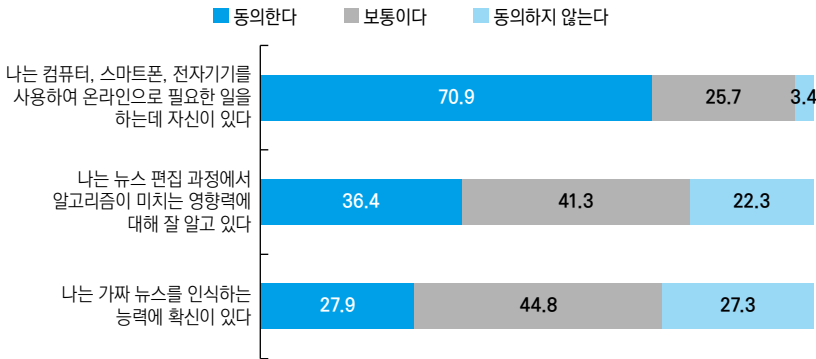
미디어 리터러시는 본래 미디어 기반의 다양한 콘텐츠를 읽고 생산하는 능력을 의미한다. 하지만 기술 발달로 인해 미디어 리터러시의 범위는 디지털 기반 정보에 대한 비판적 이해와 활용까지 확장하고 있다. 미디어 리터러시는 미디어

기술의 발달로 사회 구성원이라면 누구나 기본적으로 갖추어야 할 소양으로, 최근에는 알고리즘이 미디어 이용에 미치는 영향력이 점차 증가함에 따라 알고리즘에 관한 이용자 교육 필요성도 논의되고 있다. 본 설문 조사에 따르면 포털 뉴스 이용자 10명 중 7명이 포털 뉴스 배열에 알고리즘이 활용된다는 사실을 인지하고 있었다. 그렇지만 미디어에 유통되는 정보를 비판적으로 이해한다든지 알고리즘이 구체적으로 미치는 영향력을 파악하는 것은 또 다른 이야기이다.

일반적으로 논의되는 미디어 리터러시를 포함하여 포털 뉴스 이용자가 스스로 느끼는 리터러시 효능감은 어느 정도인지 물었다. 우선 전자기기를 활용하여 온라인으로 특정 업무를 수행하는 것에 관한 효능감은 응답자의 70.9%가 자신 있다고 답했으며, 자신 없다는 응답자는 3.4%에 지나지 않았다. 알고리즘 영향력에 관해 잘 알고 있다고 생각하는지에 대한 질문에 응답자의 36.4%가 동의한다고 답했다. 단순히 미디어 기기를 활용하는 것에 대한 효능감과 비교하면 절반 정도에 불과했다. 가장 낮은 효능감을 보인 것은 가짜 뉴스를 인식하는 능력이었다. 가짜 뉴스를 인식하는 능력에 확신이 있는지에 대한 질문에 응답자의 27.9%만이 확신이 있다고 답했으며, 27.3%는 확신이 없다고 응답했다. 이러한 결과는 다양한 접근성 교육도 중요하지만, 알고리즘과 같은 새로운 기술이 미치는 영향력을 알리고, 뉴스 전반에 대한 비판적 이해를 높이는 것이라는 점을 확인시켜준다.

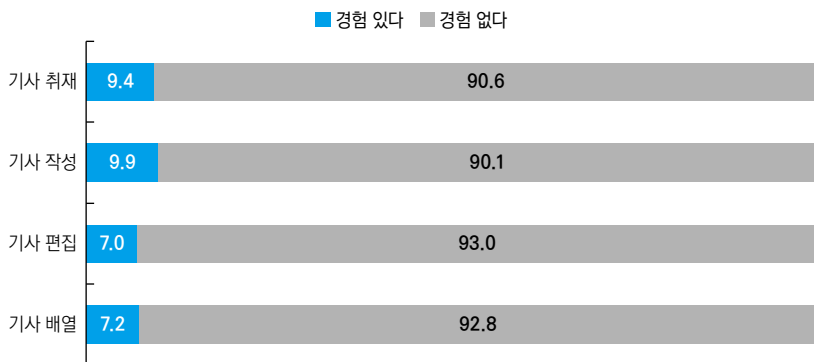
〈그림 4-40〉 미디어 리터러시에 관한 자기효능감

(단위: %, n=1,000)



〈그림 4-41〉 뉴스 생산 과정 참여경험

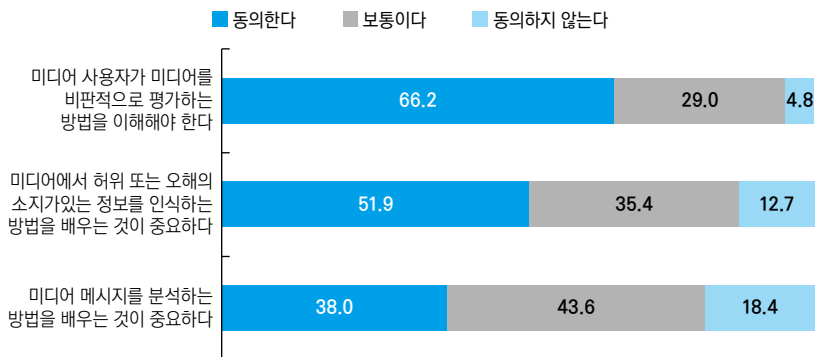
(단위: %, n=1,000)



비판적 미디어 리터러시를 높이기 위해서는 실제 뉴스 제작 과정에 대한 이해가 필요하다. 이를 확인하기 위해 뉴스 제작과 관련된 취재, 기사 작성, 편집, 배열 과정을 경험해 본 적이 있는지 물었다. 대부분의 뉴스 이용자가 뉴스 제작 과정에 참여해 본 경험이 없다고 답했다. 실제 경험을 한 응답자는 10명 중 1명 수준이었으며, 기사 취재와 작성보다는 편집과 배열을 경험해 본 응답자가 더 적었다. 기사 취재는 9.4%, 기사 작성은 가장 많은 9.9%였다. 기사 편집은 7.0%, 기사 배열은 7.2%가 경험이 있다고 답했다.

〈그림 4-42〉 미디어 리터러시 필요성

(단위: %, n=1,000)



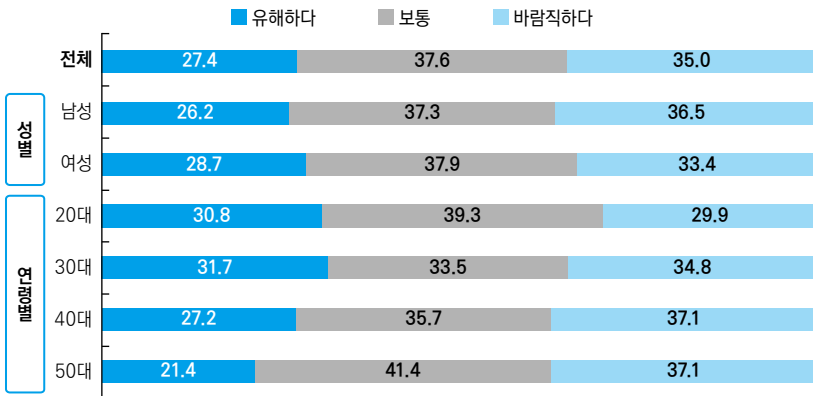
뉴스 제작 과정을 직접 경험한 사람은 소수에 불과하지만, 응답자는 미디어 리터러시 필요성에 대해 공감하는 태도를 보였다. 미디어 사용자가 미디어를 비판적으로 평가하는 방법을 이해해야 한다는 질문에 응답자의 66.2%가 동의했다. 미디어에서 허위 또는 오해의 소지가 있는 정보를 인식하는 방법을 배우는 것이 중요하다는 응답도 51.9%에 달했다. 미디어 메시지를 분석하는 방법을 배우는 것이 중요하다는 응답자는 38.0%로 나타났다. 단순히 미디어 메시지를 분석하는 것을 넘어 허위정보나 오정보를 인식하는 역량을 키우고 비판적으로 평가하는 방법을 이해하는 것이 필요하다는 점에 다수의 이용자가 공감했다.

(2) 포털 사회적 영향력과 규제 필요성

포털에 뉴스 이용자가 집중되고 포털이 뉴스 유통에서 차지하는 영향력이 높아짐에 따라 포털 규제에 관한 다양한 논의가 이루어지고 있다. 규제 필요성에 대한 조사에 앞서 포털의 뉴스 편집이 가지는 사회적 영향력에 관한 인식을 물었다. 포털의 뉴스 편집이 사회에 미치는 영향력에 대해 유해한지 바람직한지, 부정적인지 긍정적인지, 위협적인지 안정적인지 세 가지 항목으로 살펴보았다.

〈그림 4-43〉 포털 뉴스 편집이 사회적으로 미치는 영향력이 유해한가

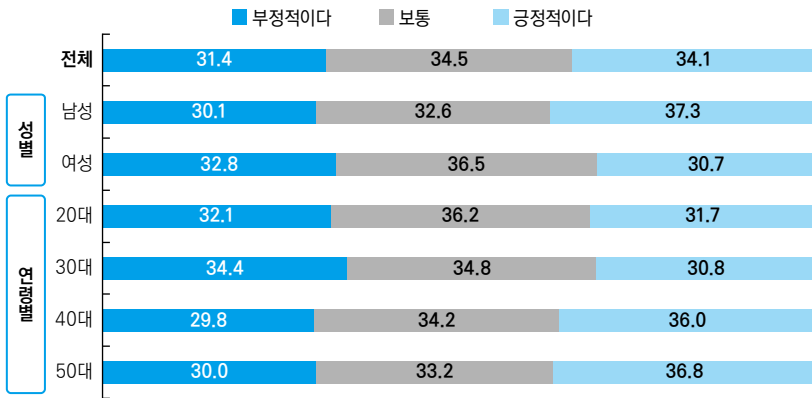
(단위: %, n=1,000)



포털 뉴스 편집이 사회적으로 해로운 영향을 미친다는 응답은 27.4%가 동의했으며, 바람직한 영향을 미친다는 답변이 35.0%로 더 높았다. 성별로 살펴보면, 여성이 남성에 비해 부정적 인식이 약간 높았으며, 연령대별로 살펴보면 20대와 30대가 40대와 50대보다 부정적 인식이 높았다. 특히 20대는 바람직한 영향을 미친다는 의견이 29.9%, 해로운 영향을 미친다는 응답이 30.8%로 부정적 의견이 긍정적 의견보다 약간 높았다.

〈그림 4-44〉 포털 뉴스 편집이 사회적으로 미치는 영향력이 부정적인가

(단위: %, n=1,000)

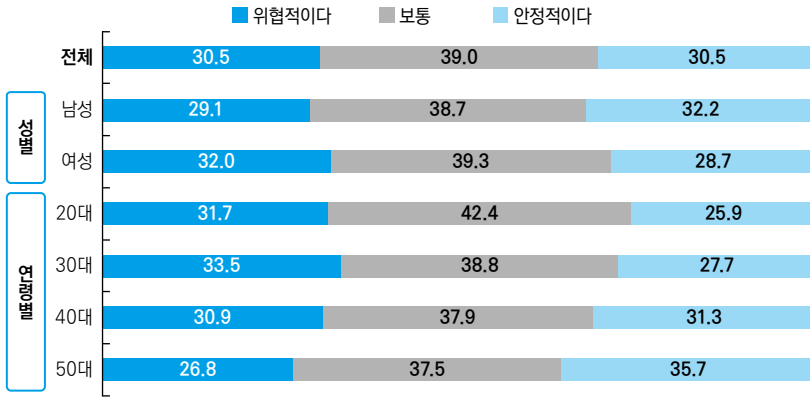


포털 뉴스 편집이 사회적으로 미치는 영향력이 부정적이라는 응답은 31.4%, 긍정적이라는 응답은 34.1%로 긍정적 평가가 약간 높았다. 성별로는 여성이 남성보다 부정적으로 인식하는 응답이 높았다. 연령대별 결과도 20대와 30대가 40대나 50대에 비해 부정 평가가 높았다. 동일하게 20대는 포털 뉴스 편집이 사회적으로 미치는 영향력이 부정적이라는 응답이 32.1%로 긍정적이라는 응답 31.7%에 비해 높았다.

포털 뉴스 편집이 사회적으로 미치는 영향력이 위협적인지를 묻는 질문에서 위협적이라는 응답자는 30.5%, 안정적이라는 응답자는 30.5%로 같은 결과를 보였다.

〈그림 4-45〉 포털 뉴스 편집이 사회적으로 미치는 영향력이 위협적인가

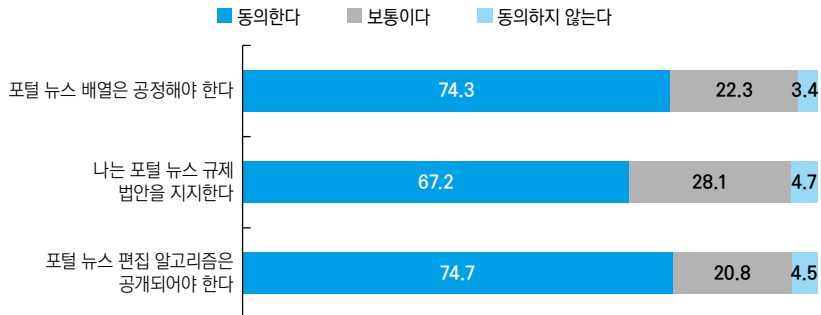
(단위: %, n=1,000)



성별 차이를 살펴보면 위협적이라는 응답이 남성보다 여성이 높았다. 연령별 차이는 50대를 제외하고 위협적이라는 응답이 높은 편인 것을 확인할 수 있었다. 위협적이라는 답변은 20대 31.7%, 30대 33.5%, 40대 30.9%, 50대 26.8%였다. 포털 뉴스 편집의 사회적 영향력 관련 질문에서 전반적으로 긍정적 응답이 소폭 높거나 비슷한 수준으로 나타났다. 성별로는 여성이 느끼는 부정적 인식이 높았으며, 연령대별로는 2030세대가 4050세대보다 높았다.

〈그림 4-46〉 포털 뉴스 배열 규제에 관한 인식

(단위: %, n=1,000)



포털 뉴스 배열이 사회적으로 미치는 영향력에 대한 부정적 인식은 긍정적 인식과 유사한 수준으로 나타났으나, 포털 뉴스가 공정해야 한다는 질문에 74.3% 응답자가 동의했다. 동의하지 않는 응답자는 3.4%로 매우 소수에 불과했다. 포털 뉴스 규제 법안을 지지하는지에 관한 질문에서는 응답자의 67.2%가 동의한다고 답변했다. 포털 뉴스 편집 알고리즘이 공개되어야 한다는 응답자도 74.7%에 달했다. 뉴스 편집 알고리즘의 공개가 가질 수 있는 문제점도 있으나 다수의 포털 뉴스 이용자는 알고리즘 공개가 필요하며 포털 뉴스 규제 법안에 지지하는 모습을 보였다.

3. 소결

본 설문조사는 현재 포털 뉴스 유통 상황과 뉴스 배열 알고리즘 적용에 대한 뉴스 이용자의 인식에서부터 시작하여 포털 뉴스 이용자가 판단하는 신뢰 요인을 탐색적으로 파악했다. 이를 통해 저널리즘 분야 전문가들이 바라보는 당위적인 원칙과 뉴스 이용자가 바라보는 개별 기사와 뉴스 배열의 신뢰도 판단에 대한 내용을 비교함으로써 향후 저널리즘 가치를 반영하는 알고리즘 설계의 기반을 제공하고자 했다. 향후 이러한 내용을 바탕으로 만들어지는 알고리즘이 실제 뉴스 유통 상황에 적용되지 않더라도, 이용자의 리터러시 교육에 활용됨으로써 신뢰 문제 해결을 위해 기여할 수 있는 영역이 무엇인지 살펴보고자 했다.

포털 뉴스 이용에 대한 설문조사 결과는 알고리즘 영역에 관한 사회적 논의가 더욱 중요해 질 것이라는 사실을 확인시켜준다. 과거 뉴스를 접하던 주요 경로였던 ‘잡지’, ‘종이신문’, ‘라디오’는 이용자가 소수일 뿐만 아니라 이용자 참여 수준도 매우 낮다는 사실을 알 수 있다. 포털이 주요한 뉴스 이용경로라는 사실은 널리 알려졌지만, ‘온라인 동영상 플랫폼’과 ‘소셜미디어’도 주요한 뉴스 이용경로로 부상하고 있다. 이러한 이용경로 모두 알고리즘이 중요한 역할을 차지하는 온라인 뉴스 유통 채널이다. 뉴스 이용자들은 정보를 얻기 위한 목적에서 포털을 이용하며, 다양한 언론사의 뉴스를 비용 지불 없이 모바일에서 빠르게 확인할 수 있기 때문에 포털 뉴스를 이용하게 된다고 답했다. 특히 모바일 이용 비율의 증가는 단순히 모바일로 인한 이용방식이 변화했음을 이야기하는 것이 아니라 제한된 화면 공간 안에서 어떤 방식으로 뉴스를 노출할 것인지에 관한 기준을 논의하는 것이 점차 중요해질 것이라는 사실을 보여준다.

포털 뉴스 이용자 10명 중 7명은 포털이 뉴스 배열에 알고리즘을 활용한다는 사실을 인지하고 있었다. 알고리즘 뉴스 배열에 관한 일반적 인식을 통해 알고리즘을 통한 뉴스 배열 결과가 상업성과 편향성을 보이며 신뢰도가 낮을 것으로 생각하고 있다는 사실을 알 수 있었다. 뉴스 배열 알고리즘으로 인한 변화에 관해 긍정적인 시각과 부정적인 시각이 함께 나타났다. 기존과 달리 다양한 뉴스

를 접할 수 있으며 중요한 뉴스를 볼 수 있게 될 것이라는 의견도 있었으나, 관심 영역에서 벗어나는 기사를 보기 어려워졌으며 다른 방식으로 뉴스를 이용하는 경향이 증가하였다는 응답도 있었다. 일반적으로 논의되는 개인정보 유출이나 원하는 정보에 대한 접근이 차단될 수 있다는 우려는 낮은 수준이었다. 알고리즘에 관한 긍정적 시각과 부정적 시각이 공존하는 가운데 신뢰와 관련한 인식에서는 사람보다 알고리즘의 우위를 확인할 수 있었다. 신뢰할 만한 뉴스 배열 주체를 물었을 때, 사람보다는 기계가 신뢰할 만하다는 응답이 65.7%로 사람보다 높았으며 다양한 평가 기준 중 심층성을 제외한 공정성, 투명성, 다양성, 객관성, 정확성, 화제성을 신뢰할 만하게 배열하는 주체로 사람보다는 기계일 것이라고 답했다.

현재 뉴스 신뢰도 조사 결과는 매우 낮은 수준을 보였다. 뉴스 전반에 대해 신뢰한다는 응답은 25.8%였으나, 기사를 신뢰한다는 응답은 16.5%, 언론사를 신뢰한다는 응답은 13.3%, 포털은 기자와 언론사보다는 상대적으로 높은 20.9%로 나타났다. 전반적으로 뉴스에 관한 신뢰도가 낮은 상황에서 전체 설문 응답자(1,000명) 기준으로 살펴보면, 뉴스 신뢰도 평가에 동의하는 응답자는 86.7%였다. 뉴스 신뢰도 기계적 평가에 동의하는 응답자는 전체 응답자 기준 63.5%였다. 찬성의 주요 이유는 사람이 가진 편견을 제거할 수 있기 때문이라고 답했지만, 반대의 주요 이유는 모든 기사를 같은 기준으로 판단하는 것이 어렵기 때문이라고 답했다. 알고리즘 기반 신뢰도 평가 점수를 뉴스 배열 기준에 포함하는 것에 동의하는 응답자는 60.5%였다. 찬성의 주요 이유는 신뢰도를 중심으로 다양한 뉴스를 볼 수 있기 때문이었으며, 반대의 주요 이유는 신뢰도 평가를 (포털 배열 기준으로 포함할 만큼) 믿을 수 없기 때문이라는 응답이 많았다. 이를 통해 뉴스 신뢰도가 낮은 상황에서 뉴스에 대한 신뢰도 평가가 필요하다는 응답이 다수인 것을 확인할 수 있었다. 하지만 알고리즘에 관한 신뢰도는 아직은 상대적으로 높지 않으며, 기계적 신뢰도 평가에 동의한 응답자 대부분이 포털 뉴스 배열 기준으로 알고리즘 신뢰도 평가 점수를 활용하는 것에 동의했음에도 정확성에 대한 우려로 포털 배열 기준에 활용되는 것에 부정적 태도를 나타낸 사례도 있었다.

실제 뉴스 신뢰도 평가에 영향을 미치는 다양한 요인의 중요성을 포털 뉴스

이용자에게 물어본 결과에 따르면, 가장 중요한 것은 기사 자체 요인이었다. 기사 요인 중에서 출처 확인이나 사실 검증 관련된 기사의 객관성, 중립적인 보도와 관련된 공정성, 협찬 기사의 명확한 표현과 같은 투명성이 뉴스 신뢰도를 판단에 중요하다고 답했다. 기자 요인도 어느 정도 중요성을 가지고 있었다. 기자가 오보일 때 기사를 수정하고 공개하는지와 일반 대중이 생각하는 문제를 이해하고 있는지가 신뢰도 평가에 중요하다는 의견이 있었다. 이러한 응답은 기자가 대중이 생각하는 문제의식과 거리가 먼 기사를 작성하는 경향이 있다는 뉴스 이용자의 인식을 반영한 것일 수 있다. 언론사 요인 중에서는 언론사 과거 기록이 신뢰도 판단에 영향을 주었다. 과거 오보 여부나 허위보도 작성 여부가 신뢰도 판단에 영향을 미친다고 보았다. 뉴스 이용자는 단순 실수라고 하더라도 기자나 언론사의 오보가 반복되었다는 평가를 통해 신뢰도를 판단하는 데 활용한다고 볼 수 있었다.

이러한 이용자의 신뢰도 평가 요인 중에서 전문가들이 중요하다고 평가하는 요인이지만 뉴스 신뢰도 평가에서 중요성이 낮다고 언급한 요인도 있었다. 전문가들은 다른 언론사 기사를 받아쓴 보도가 아닌 자체 생산 기사에 높은 평가를 했으나 이용자 시각에서는 중요성이 낮다고 언급했다. 언론사 소유 구조나 수익 모델은 전문가 관점에서 저널리즘의 신뢰성에 영향을 미칠 수 있는 중요한 요소였으나 이러한 요인 역시 이용자 입장에서는 중요성이 낮다는 평가를 받았다. 물론 이러한 요인은 일반 뉴스 이용자 입장에서는 파악하기 어려운 부분이기 때문에 개인이 기사 신뢰도를 평가할 때 활용되지 못하는 측면도 있다. 비슷한 이용자가 본 항목을 추천하는 방식은 추천 알고리즘에서 주로 활용되지만, 이용자들은 많은 이용자가 보거나 자신이 신뢰하는 사람이 보는 뉴스라고 신뢰하지는 않는 것으로 나타났다. 다만, 신뢰를 평가하는 요소와는 전혀 관련이 없는 특정 주제에 관해 자신과 동의하는 보도를 한다든지, 언론사 기자가 나와 같은 정치적 견해를 가지고 있는지가 뉴스 신뢰도 평가에 영향을 미치는 중요한 요소라고 답한 응답자도 각각 38.8%, 29.3%로 나타나 이용자 개인이 판단하는 뉴스 신뢰 지표에 대해 고민할 여지가 있다는 사실도 확인할 수 있었다.

뉴스에 다양한 신뢰 지표를 포함하도록 유도하는 방식이 뉴스에 대한 신뢰도를 얼마나 증가시킬 수 있는지를 탐색적으로 살펴보기 위한 실험에서, 개별 신뢰 지표로 사용한 기사 작성시간, 기사 수정시간, 기사 수정 내역, 기사 분류정보, 취재원 정보, 기자 사진, 기자 이메일, 기자 이력은 언론사의 일반적 평가에는 큰 영향을 주지 않았다. 다만 언론사 취재방식과 기자 개인에 대한 평가에서는 통계적으로 유의한 영향을 미쳤다. 뉴스 평가 알고리즘은 개별 언론사에게 저널리즘 영역에서 이루어지던 좋지 않은 관행을 개선하는 방식으로도 영향을 미칠 수 있다. 이러한 요소가 포함되지 않은 기사에 대해 낮은 평가를 부여하는 알고리즘은 언론사의 기사 생산 방식의 변화를 유도할 수 있다. 언론사 신뢰도와 같은 평가에서는 즉각적인 영향을 미치지 못했고 언론사 만족도를 바로 높이지는 못했으나, 장기적으로는 언론사 취재 방식과 기자 개인에 대한 긍정적 평가를 통해 신뢰도를 개선할 수 있는 가능성이 있을 수 있다.

기사 실험에서 확인할 수 있는 사실 중 하나는 뉴스 이용자가 생각하는 개별 신뢰 요인의 중요성과 기사를 볼 때 인지하는 요인에서 차이가 크다는 점이다. 기사 수정 내역을 공개하고 취재원을 밝히는 것은 뉴스 신뢰도에 영향을 미치는 중요한 요인이라고 생각했으나, 기사를 볼 때 해당 내용을 인지하는 응답자는 많지 않았다. 오히려 중요성은 낮지만 기자 사진을 인지하는 응답자가 가장 많았다. 실존하지 않는 기자의 가상 프로필을 만드는 방식이 가짜 뉴스를 전파하는데 사용된 사례가 있다(Rawnsley, 2020). 가상의 프로필을 위해 딥페이크를 통한 인공지능 생성 이미지를 활용했다. 이를 통해 단순히 기자 사진의 유무가 아닌, 확인할 수 있는 출처를 검증하는 것이 저널리즘 신뢰도 측면에서 중요성을 가질 수 있다는 사실도 확인할 수 있었다. 뉴스 트러스트 위원회에서는 신뢰도 평가 알고리즘에서 현실적으로 활용할 수 있는 요소로 기자 이메일에 중요한 가중치를 두었다. 하지만 설문 조사 결과에서 응답자들이 기자 이메일을 인식하는 정도는 가장 낮았다. 이러한 결과는 개별 요소의 포함 여부보다는 알고리즘을 바탕으로 이용자가 확실히 신뢰할 수 있는 검증된 기사인지를 표시해주는 마크나 배지를 활용하는 방식도 고려해볼 수 있다는 사실을 보여준다.

배열에 관한 신뢰도 평가 실험에서는 단순히 이러한 신뢰 표식을 활용하는 방식보다는 점수를 함께 포함하는 방식이 신뢰도가 더 높다는 사실을 확인할 수 있었다. 뉴스 배열 결과만으로는 뉴스 이용자가 어떤 기준으로 배열이 되었는지 확인하기 어려워 공정성, 다양성, 정확성, 심층성, 투명성, 객관성, 화제성과 같은 다른 기준에서 인지된 차이를 만들어내지는 못했다. 하지만 신뢰 측면에서는 통계적으로 유의한 차이를 나타낸 것을 볼 수 있었다. 언론사명 포함 여부와 배열방식의 영향력을 살펴본 실험에서 언론사명 포함 여부는 배열에 관한 신뢰도에서 영향을 미치지 못했다. 이러한 결과는 언론사명이 실제 언론사가 아닌 가상의 언론사였기 때문일 수도 있고, 포털에서 뉴스를 접할 때 뉴스 이용자가 출처를 인지하지 못하는 출처 망각에 의한 효과일 수도 있다. 모두 다른 언론사명이 포함되어 있었기 때문에 동일한 이름을 가진 언론사가 다수 포함되었을 경우와 그렇지 않은 경우 효과는 다르게 나타날 수 있다. 배열방식의 경우 신뢰도에 영향을 주지는 못했지만 뉴스 배열이 공정한지에 대한 부분에서는 이용자 인식에 영향을 주었다. 높은 신뢰도 점수 순서로 배열한 뉴스 기사를 보고 이용자는 뉴스 배열이 공정하다는 느낌을 받은 것으로 나타났다.

포털 뉴스 이용자는 미디어 접근성과 관련된 리터러시에서는 다수가 효능감을 나타냈으나, 뉴스 편집 과정에서 알고리즘이 미치는 영향력이라든지 가짜 뉴스를 인식하는 능력에 관해서는 낮은 효능감을 보였다. 이러한 부분을 이용자가 인지하기 위해서는 실제 뉴스 생산 과정에 대한 리터러시 교육이 선행되어야 한다. 실제 뉴스 생산 과정에 어느 정도라도 참여해 본 응답자는 10명 중 1명 수준으로 매우 낮았다. 다수 응답자는 미디어를 비판적으로 평가하고 허위정보를 인식하는 방법을 배우는 것이 중요하다는 점에 공감했다. 뉴스 트러스트 위원회의 알고리즘은 뉴스 신뢰도에 대한 문제의식에서 출발하여 저널리즘 가치를 포함하는 알고리즘을 만들기 위한 시도였다. 실제 뉴스 기사 평가나 배열에서 활용되지 않더라도, 이를 활용한 뉴스 리터러시 또는 알고리즘 리터러시 교육은 이용자의 리터러시 역량을 키우는 방식으로 활용될 수 있다. 포털 뉴스 편집이 사회적으로 미치는 영향에 대해서는 긍정과 부정 의견이 비슷한 수준으로 나타났지만, 다수

의 응답자가 포털 뉴스 편집은 공정해야 한다는 점에 동의했으며, 알고리즘이 공개될 필요성이 있다고 답했다.

알고리즘이 미디어 영역에 미치는 파급력이 확대됨에 따라 알고리즘에 관한 다양한 논쟁이 지속되고 있다. 초기 알고리즘에 관한 논쟁은 알고리즘은 실제 데이터를 반영하는 투명한 것이기 때문에 편향을 완전히 제거하는 것은 불가능하다는 견해를 가졌다. 하지만 알고리즘도 현실을 재현하는 일종의 미디어로 어떤 방식으로 사회적 실재를 보여줄 것인지에 대한 논의가 필요하다. 즉, 알고리즘은 무엇이 옳은지 그른지에 대한 논의를 통해 사회적 합의가 이루어져야 하는 공간이다. 뉴스 트러스트 위원회의 시도는 이러한 문제 제기를 위한 시도였다. 본 설문조사는 뉴스와 관련된 알고리즘에 대한 이용자 인식을 바탕으로 이러한 논의에서 고려할 수 있는 내용을 제시했다. 신뢰를 판단할 수 있는 다양한 요인이 있지만, 우리가 어떤 것을 신뢰할 만하다고 판단할지에 대한 부분도 사회적 합의가 이루어져야 하는 영역이다. 설문조사 결과에서 나타난 것처럼 자신의 의견과 일치하는 뉴스나 기자가 같은 정치적 견해를 가졌는지를 신뢰도 판단에 활용하는 뉴스 이용자도 있다. 전문가가 생각하는 부분과 일반 이용자가 판단하는 신뢰에 대한 차이도 존재한다. 이러한 탐색적인 시도가 기존에 명확하게 정의되지 않았던 다양한 근본적 개념에 대해 다시 논의할 수 있는 계기를 마련할 수 있을 것이다.

05

결론 및 제언

뉴스 트러스트 위원회는 이용자 선호에 기반한 알고리즘과 달리 품질 높은 기사를 판별하는 기준에 근거한 알고리즘을 만들고자 하는 시도였다. 프로젝트가 종료한 이후 알고리즘을 구성하는 규칙은 모두 투명하게 공개하였다. 일부에서는 공개한 알고리즘을 보고 무슨 근거로 품질 높은 기사를 판별하는 기준을 만들었는지 의문을 가질 수 있다. 제한적인 계량 요인만으로 기사가 가지는 다양한 측면을 모두 반영하는 알고리즘을 만들기는 쉽지 않은 일이었다. 프로젝트는 3년간 이루어졌지만, 저널리즘 가치에 기반하여 정교하게 기사를 판별해내는 알고리즘을 만들었다고 이야기하기는 어렵다. 프로젝트의 목적은 저널리즘 가치를 반영하는 알고리즘을 만들고자 하는 것이었다. 이런 측면에서 뉴스 트러스트 위원회 결과물이 가지는 함의는 알고리즘에 관한 대안적인 시각을 제시하고 현재 온라인에 유통되고 있는 기사가 지향해야 할 방향이 무엇인지 논의했다는 점에 있다. 이를 통해 알고리즘 리터러시에 대한 필요성을 끌어냄과 동시에 바람직한 품질 높은 뉴스의 기준이 무엇인지 사회적인 합의가 필요함을 이야기했다.

알고리즘에 대한 오해 중 하나는 알고리즘이 데이터 기반으로 학습한 결과를 그대로 보여주는 투명한 도구라는 생각이다. 이러한 생각은 알고리즘이 발생시키는 편향의 결과를 이용자 행동을 그대로 반영하는 것일 뿐이라는 생각으로 이끈다. 이러한 결론은 일부는 사실이지만 일부는 사실이 아니다. 다수의 알고리즘은 실제로 데이터에 기반하여 가장 적합한 결과를 추정해내는 기술을 활용한다.

하지만 알고리즘이 그렇다고 하더라도 알고리즘이 발생시키는 편향에서 자유로운 것은 아니다. 최근 추세는 알고리즘으로 인해 발생하는 편향을 최소화하고 사회적으로 바람직한 결과를 제시하는 방안도 연구되고 있다. 예를 들면, 단어 빈도와 같은 말뭉치 규칙성이 성별 편향을 증폭시키는 부정적 결과를 낳을 수 있기에, 이러한 편향을 완화하기 위한 조정작업이 필요하다는 사례도 있다 (Wang, Lin, & Rajani, 2020). 알고리즘은 미디어 영역에서도 점차 강력한 영향력을 미치며 사람들의 일상을 변화시킨다. 단순히 이용자 행동의 반영이라고 넘길 것이 아니라 세계를 어떤 방식으로 드러내야 하는지 적극적인 고민과 사회적 합의가 필요한 공간이다.

프로젝트의 주요 성과 중 하나는 온라인 뉴스 품질에 대한 규범적 논의를 위한 시작점을 제시했다는 것이다. 프로젝트가 처음 시작되었을 때는 위원들 사이에서도 프로젝트의 목적과 방향에 관한 오해가 있었다. 초기에는 정확도가 떨어지는 결과물이 나올 수밖에 없는 상황에서 프로젝트를 진행하는 것에 회의적인 의견을 표명한 위원도 있었다. 이용자 데이터에 의존하지 않는 설명 가능한 대안적인 알고리즘을 제시하기 위해 품질 높은 기사에 대한 합의를 만들어내는 것이 필요함에도 이용자 데이터 활용에 대한 논의는 반복적으로 이어졌다. 이러한 종류의 시도는 모험적이고 낯선 것이었기 때문이다. 기계 학습에 기반한 새로운 알고리즘이 존재하는데도 규칙 기반의 오래된 알고리즘을 만들어야만 하는 이유에 대해 공감대를 형성하기까지는 오랜 시간이 걸렸다. 규칙 기반의 알고리즘은 낡고 오래된 방식이라는 인식 때문이기도 했다. 또한, 기계 학습이 우리가 가지고 있는 모든 문제를 마술처럼 해결해 줄 것이라는 잘못된 시각도 있었다. 하지만 기계 학습은 데이터에 기반하여 인간의 의사 결정을 유사하게 모사할 뿐이다. 우리가 원하는 바람직한 결과를 위해서는 우리가 원하는 결과를 명확하게 합의하고 만들어내는 과정이 먼저 이루어져야만 했다. 프로젝트 초반에 이용자 데이터에 기반한 학습을 통해 알고리즘을 만들어야 하는 것은 아닌지 반복적인 논쟁이 이루어진 것은 이러한 오해에서 비롯했다고 볼 수 있다.

논의의 진척이 쉽지 않았던 이유는 지나치게 막연한 작업이었기 때문이기도 하다. 저널리즘에서 어떤 기사를 품질 높은 기사로 평가하는지에 대한 논의는 오래전부터 논의되었던 주제이다. 하지만 기계가 이해할 수 있는 수준에서 품질 높은 기사가 무엇인지 정의하는 문제는 추상적으로 이루어지던 논의와는 달랐다. 모두가 품질 높은 기사가 무엇인지 알고 있다고 생각했지만 다양한 위원들의 논의과정에서 우리가 품질 높은 기사를 구체적으로 정의하는 것이 가능한지에 대한 의문으로 이어졌다. 예를 들면, 저널리즘에서 가장 오랫동안 언급되었던 균형성이라는 개념을 어떻게 정의할 것인지에 관해서도 의견이 나뉘는 모습을 발견할 수 있었다. 과연 현재 대한민국의 현실에서 균형성이란 바람직한 저널리즘 가치인지에 대한 논의도 이어졌다. 오히려 기계적인 균형성이 나쁜 저널리즘을 이끄는 것은 아닌지에 대한 회의도 있었다. 기계가 측정을 통해 균형성을 판단하기 위해서는 균형성이란 무엇인지에 대한 개념적 합의가 우선해야 했음에도 불구하고 저널리즘에서 오랫동안 논의되어 온 균형성을 현재 시점에서 무엇인지 정의하는 작업도 쉬운 것이 아니었다. 취재원 숫자가 고르게 분포되면 균형성을 갖춘 것인지, 인용한 내용을 가지고 판단이 이루어져야 하는지, 균형성 자체가 모든 기사에 필요한 것인지에 대한 것을 합의해나가는 과정도 쉽지 않았다.

우리는 개념에 대해 이해하고 있다고 생각하지만, 시간의 변화는 우리가 알고 있던 개념도 새롭게 정의할 것을 요구한다. 저널리즘에 관한 원칙은 오래전부터 논의되었지만, 온라인 뉴스에서 저널리즘 원칙은 무엇이 되어야 하는지는 어떤 의미에서 제한적이었다. 전통적으로 바라보던 저널리즘 원칙을 개념적인 수준에서 시작했던 논의는 계량 요인과 저널리즘 가치를 연결하는 과정을 거치며 점차 구체화하였다. 이러한 과정도 간단하지 않았다. 저널리즘 분야의 전문가로 구성된 위원회였으나 위원들도 기초적인 수준의 계량 요인을 추상적인 수준의 저널리즘 가치와 연결하는 작업에서 확신하기 어려웠다. 인간이 직관적으로 하는 판단을 기계적 언어로 번역하는 것이 막연한 시도였기 때문이다. 이로 인해, 제한적인 요인을 바탕으로 기준을 만드는 것이 가능한지 회의적이었던 위원들은 정확성에

대한 우려로 기사 외적인 평가 요인을 도입할 필요성에 관해 언급하기도 했다.

기준을 정하기 위한 과정에서 다양한 논쟁이 있었으나 실제 온라인에서 유통되는 표본 기사를 확인하고 평가하면서 우리가 온라인에서 지향해야 하는 기사가 무엇인지 점차 구체화하는 과정을 거쳤다. 실제 우리가 하루에 온라인에서 보는 기사는 온라인에 유통되고 있는 기사의 1%에도 미치지 못한다. 오랜 기간 저널리즘을 연구한 학자와 현업에서 종사한 기자도 기사를 직접 읽으며 새로운 경험을 하게 되었다고 이야기했다. 반복적인 논쟁과 합의를 통해 프로젝트 목표를 향해 조금씩 나아갔지만, 모두가 합의할 수 있는 품질 높은 기준을 마련하기는 쉽지 않은 일이었다. 부정적인 것이 무엇인지에 관한 합의는 상대적으로 쉬운 편이었다. 온라인 저널리즘에서 발생한 문제점을 해결하는 데 필요한 것이 무엇인지는 합의할 수 있었다. 하지만 좋은 저널리즘이 무엇인지는 위원 개인마다 관점의 차이가 존재했다. 프로젝트는 저널리즘 가치 기반으로 다양한 기사 배열을 뉴스 이용자가 직접 선택할 수 있는 방식으로 알고리즘을 제시했다. 대안적인 알고리즘이라는 목표에 충실한 해결책이기도 했다.

3년에 걸쳐 이루어진 개발과정과 논의를 모두 전달하기는 어려운 일이지만, 저널리즘 가치 기반 알고리즘을 만들기 위한 과정에서 어떤 주요한 논쟁이 발생했는지를 전달하고자 했다. 일부 사람들은 여전히 판별 기준이 자의적이며 이런 기준으로 뉴스 기사를 판별하는 알고리즘이 적절하지 않다고 생각할 수 있다. 3년이라는 기간 동안 정확도가 높은 정교한 알고리즘을 만들어내는 것은 기술적으로 제한이 있었다. 시간이 더 주어진다고 해도 가능하지 않을 수도 있다. 저널리즘 가치 기반의 알고리즘을 만드는 것의 목적은 온라인 저널리즘에서 과연 어떤 것이 바람직한 기사인지 사회적 합의가 필요함을 알리는 신호이기도 했다. 플랫폼 기업은 이용자 데이터에 기반한 알고리즘을 통해 우리가 파악하기 어려운 결과물을 제시한다. 하지만 뉴스 트러스트 프로젝트에서 제시한 내용은 알고리즘이 단순히 이용자의 선호를 그대로 반영하는 공간이 아닌 사회적 합의를 바탕으로 무엇을 보여주어야 하는지 논쟁이 필요한 공간이라는 점을 시사해준다. 사회

적으로 바람직한 뉴스가 무엇인지에 대한 논의는 이제 기초적인 단계가 시작되었을 뿐이다. 이에 관한 합의를 통해 올바른 알고리즘이 보여주어야 하는 현실이 무엇인지에 관한 논의가 필요한 시점이다.

참고문헌

- 김대원·윤영민 (2015). SNS에서 형성된 신뢰가 위기 시 방어막이 될 수 있는가. 『한국언론학보』, 59권 2호, 196~225쪽.
- 김영옥·진민정·강신규 (2014). 『저널리즘의 품질: 평가 기준과 모델』. 서울: 한국언론진흥재단.
- 김위근·안수찬·백영민 (2018). 『한국의 언론 신뢰도: 진단과 처방』. 서울: 한국언론진흥재단.
- 민 영 (2016). 신뢰의 조건: 저널리즘 전문성과 정파적 편향성이 언론 신뢰와 정치 신뢰에 미치는 영향. 『한국언론학보』 60권 6호, 127~156쪽.
- 박승택·성인재·서상원·황지수·노지성·김대원 (2017). 기계 학습 기반의 뉴스 추천 서비스 구조와 그 효과에 대한 고찰: 카카오의 루빅스를 중심으로. 『사이버커뮤니케이션학보』 34권 1호, 5~48쪽.
- 박이란·이소은 (2020). 「디지털 뉴스 리포트 2020 한국」. 한국언론진흥재단-로이터저널리즘 연구소.
- 박재영·이완수 (2010). 「뉴스평가지수의 개발과 적용」. 서울: 한국언론진흥재단.
- 오세욱 (2016). 저널리즘과 알고리즘의 융합에 대한 탐색적 연구. 『사이버커뮤니케이션학보』 33권 3호, 51~101쪽.
- 이건호 (2008). 한·미 신문 기사의 심층성과 신뢰도 및 독창성 분석: 6개 한국 신문과 2개 미국 신문 1면 기사를 중심으로. 『한국언론학보』 52권 5호, 107~129쪽.
- 이건호·정완규 (2008). 한국과 미국 신문의 1면 기사 비교: 취재 영역 및 보도 형태별 취재원 출현에 따른 심층성 분석. 『한국언론학보』 52권 4호, 25~49쪽.
- 이민정·이건호 (2014). 중앙 일간지 탐사보도 특성 연구: 교육 관련 기획시리즈의 독창성 및 심층성을 중심으로. 『한국언론학보』 58권 6호, 64~89쪽.
- 이완수·박재영 (2013). 방송뉴스의 언어와 표현. 『방송과 커뮤니케이션』 14권 1호, 5~46쪽.
- 이준웅·김경모 (2008). ‘바람직한 뉴스’의 구성조건: 공정성, 타당성, 진정성. 『방송연구』, 9~44쪽.
- 황용석·김기태 (2019). 개인화 서비스 진전에 따른 자동추천 시스템 연구 동향과 방법론적 특성 연구. 『사이버커뮤니케이션학보』 36권 2호, 221~253쪽.

- Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-based systems*, 46, 109–132.
- Bogart, L. (1989). *Press and Public: Who Reads What, When, Where, and Why in American Newspapers*. Lawrence Erlbaum and Associates.
- Cremonesi, P., Garzotto, F., & Turrin, R. (2012). Investigating the persuasion potential of recommender systems from a quality perspective: An empirical study. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 2(2), 1–41.
- Curtiss, M., Bharat, K., & Schmitt, M. (2013). *Systems and methods for improving the ranking of news articles*. Google Patents. Available at: <http://www.google.com/patents/US9037575>
- De Keyser, J., & Raeymaeckers, K. (2012). The printed rise of the common man: How Web 2.0 has changed the representation of ordinary people in newspapers. *Journalism Studies*, 13(5–6), 825–835.
- De Maeyer, J. (2012). The journalistic hyperlink: Prescriptive discourses about linking in online news. *Journalism Practice*, 6(5–6), 692–701.
- Entman, R. M. (2005). The nature and source of news. In G. Overholser & K. H. Jamieson (Eds.), *The institutions of American democracy: The press* (pp. 43–65). New York: Oxford University Press.
- Gladney, G. A. (1996). How editors and readers rank and rate the importance of eighteen traditional standards of newspaper excellence. *Journalism and Mass Communication Quarterly*, Summer 1996, 319–331.
- Hermida, A. (2012). Tweets and truth: Journalism as a discipline of collaborative verification. *Journalism Practice*, 6(5–6), 659–668.
- Kovach, B. & Rosenstiel, T. (2007). *The Elements of Journalism: What Newspeople Should Know and the Public Should Expect*. Crown; 이재경(역)(2009). <저널리즘의 기본요소 - 기자가 알아야 할 것과 독자가 기대하는 것>. 한국언론진흥재단.
- Levy, S. (2010). How Google's Algorithm Rules the Web. *WIRED*, 18.03. Available at: http://www.wired.com/magazine/2010/02/ff_google_algorithm/
- Lotan, G. (2014). Networked audiences: Attention and data-informed. In K. McBride, & T. Rosenstiel, (Eds.), *The New Ethics of Journalism: Principles for the 21st Century*. Thousand Oaks, CA: CQ Press, Sage. (pp. 105–122)
- McQuail, D. (1992). *Media Performance, Mass Communication and the Public Interest*. Sage.
- Mieth, D. (1997). The Basic Norm of Truthfulness. In Clifford, C. & Traber,

- M. (ed.), *Communication Ethics and Universal Values*, Thousand Oaks, 87~104.
- Park, K., Lee, J. & Choi, J. (2017). Deep Neural Networks for News Recommendations. In *Proceedings of CIKM'17*, ACM, New York, NY, USA, 4 pages. Available at: <https://doi.org/10.1145/3132847.3133154>
- Rawnsley, A. (2020, July 7). Conservative sites like Newsmax and Washington Examiner have published Middle East hot takes from “experts” who are actually fake personas pushing propaganda, *Daily Beast*, Available at: <https://www.the-dailybeast.com/right-wing-media-outlets-duped-by-a-middle-east-propaganda-campaign>
- Toff, B., Badrinathan, S., Mont'Alverne, C., Arguedas, R, A., Fletcher, R. & Nielsen, K, R. (2020). *What we think we know and what we want to know: perspectives on trust in news in a changing world*, London: The Reuters Institute for the Study of Journalism. Available at: <https://reutersinstitute.politics.ox.ac.uk/what-we-think-we-know-and-what-we-want-know-perspectives-trust-news-changing-world>
- Verweij, P. (2012). Twitter links between politicians and journalists, *Journalism Practice*, 6(5-6), 680-691.
- Wang, T., Lin, V., & Rajani, N. (2020). Double hard-debias: Tailoring word embeddings for gender bias mitigation, *Salesforce Blog*. Available at: <https://blog.einstein.ai/double-hard-debias-tailoring-word-embeddings-for-gender-bias-mitigation/>

뉴스 기사 신뢰도
측정 방안:
뉴스 트러스트
사례를 중심으로

15,000원

한국언론진흥재단 
Korea Press Foundation

